A 3 Flavour Joint Near and Far Detector Neutrino Oscillation Analysis at T2K

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Richard Graham Calland

Friday 5th September, 2014

UNIVERSITY OF LIVERPOOL

Abstract

The Tokai-to-Kamioka (T2K) experiment is a second generation long-baseline neutrino experiment and the first to use an off-axis neutrino beam to produce narrow neutrino energy spectrum. T2K was designed to measure with precision the atmospheric mixing parameters, and also look for evidence of non-zero θ_{13}. T2K's near detector (ND280) provides constraints on the beam flux and neutrino cross-section uncertainties, as well as making valuable cross-section measurements.

This thesis describes an oscillation analysis that uses samples from both near and far detectors. Importantly, ν_μ and ν_e samples at the far detector are combined to produce a joint oscillation analysis. A Markov chain Monte Carlo is used to construct the Bayesian posterior distribution by sampling a likelihood function. From the Bayesian posterior distribution, the oscillation parameters of interest and their errors are estimated.

When analysing with only T2K run 1-4 data, the best fit point for the oscillation parameters is $\Delta m^2_{32} = (-2.57 \pm 0.11) \times 10^{-3}$ eV^2, $\sin^2\theta_{23} = 0.520^{+0.045}_{-0.050}$, and $\sin^2\theta_{13} = 0.0454^{+0.011}_{-0.014}$ at $\delta_{cp} = 0$, with the negative sign of the mass indicating that this is in the inverted hierarchy. When using the reactor experiment prior for θ_{13}, the best fit point is $\Delta m^2_{32} = (2.51 \pm 0.11) \times 10^{-3}$ eV^2, $\sin^2\theta_{23} = 0.528^{+0.055}_{-0.038}$, and $\delta_{cp} = -1.587$, and is located in the normal hierarchy. The 90% credible interval for δ_{cp} excludes 0.14–0.87, in units of π.

Declaration

I declare that the work presented in this thesis is my own. Work done by others has been clearly referenced accordingly, which includes passages of text, figures and images. This work has not been submitted for another qualification to this or any other university, with the exception of parts of Chapter 3, which started life as my Masters Dissertation [1], but has since been heavily re-worked.

Chapters 1 and 2 describe the theory motivating this thesis, along with the T2K experiment whose data was analysed. This is my summary of the theoretical work of others, and all relevant papers and documents have been referenced.

Chapter 3 discusses the construction of the Barrel Electromagnetic Calorimeter for the ND280 detector, along with its operation. I was directly involved with the construction of this detector, and its on-site operations in J-PARC. Therefore, many parts of this chapter is unreferenced as it is a personal account of the construction process.

Chapters 4 and 5 describe the development of the analysis framework used to perform the analysis detailed in Chapter 7, to which I made significant contributions as one of three developers. My main contributions to the work presented in these chapters are the development of the modular analysis framework, the GPU event-by-event reweighting systems, the implementation of far detector samples and systematic uncertainties and the kernel density estimator for parameter inference. The work in Chapter 5 was published in JINST [2].

Richard Calland

Acknowledgements

The completion of this thesis would not be possible without the help and support of many people. Firstly, I would like to thank my supervisor Christos Touramanis for having unwavering faith in my ability, and for giving me the opportunities that have lead me to this point. Of course, many thanks go to the rest of the Liverpool neutrino group members past and present.

Aside from the UK, I would like to thank all those at J-PARC in Japan, where I spent half of my time as a PhD student. The experience and friendship I gained there will stay with me forever.

Finally, I would like to thank my friends and family in the non-physics world, especially my loving wife 雅子.

I would like to acknowledge the work presented here made use of the EMERALD High Performance Computing facility provided via the Centre for Innovation (CfI). The CfI is formed from the universities of Bristol, Oxford, Southampton and UCL in partnership with STFC Rutherford Appleton Laboratory.

Contents

Chapter 1.

Neutrino Oscillation

1.1. Introduction

Neutrinos are part of the Standard Model (SM) of particle physics, however some of their behaviour is still unexplained. Most importantly, both left- and right-handed neutrino states are required to generate mass, however neutrinos have only been observed in left-handed states [3]. This means that the SM assumes neutrinos to be massless, however experimental results yield strong evidence that neutrinos are massive and mixed [4] [5]. This is a strong hint at physics beyond the SM, and makes studying the physics of the neutrino an exciting endeavour. Neutrinos offer the opportunity to probe new physics and help to understand fundamental mysteries such as why the universe is composed of non-equal parts matter and anti-matter.

This chapter forgoes the SM physics of the neutrino, as it is readily available in other resources [6]. Instead, the mathematical formalism of three generation mixing, a two flavour approximation, matter effects and an overview of the experimental results that led to the conclusion of neutrino oscillations will be presented.

1.2. Neutrino Oscillations

The mathematical prescription of describing the three flavour oscillation phenomenon will now be discussed, along with a simplified two flavour approximation and matter effects.

1.2.1. Mixing Formalism

The existence of neutrino oscillations requires that neutrinos have mass. Furthermore, the mass states must not be degenerate; it is the difference in mass of each eigenstate that drives oscillation. Here we will discuss the formalism of three neutrino mixing using the plane wave approximation to propagate the neutrino states.

We define three neutrino mass eigenstates ν_1, ν_2 and ν_3. The mass and flavour eigenstates are not decoupled; the flavour eigenstates can be expressed as a superposition of the mass eigenstates,

$$|\nu_i\rangle = \sum_{\alpha} U_{\alpha i}|\nu_\alpha\rangle, \tag{1.1}$$

where $U_{\alpha i}$ is the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) mixing matrix [7] given by

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.2}$$

where s_{ij}, c_{ij} are shorthand for $\sin(\theta_{ij})$, $\cos(\theta_{ij})$, and $e^{\pm i\delta}$ represents the complex charge-parity (CP) violating phase. It is useful to separate the PMNS matrix into three terms:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.3}$$

where the three mixing angles form separate rotation matrices. Solar and atmospheric neutrino experiments have told us that θ_{12} and θ_{23} are large (see Section 1.3), and now that θ_{13} has been confirmed to be non-zero [8] [9], the complex CP violating phase δ has the possibility to be measured.

By taking the Hermitian conjugate of the mixing matrix, the mass state can be expressed as a superposition of flavour states,

$$|\nu_\alpha\rangle = \sum_{k} U^*_{\alpha k}|\nu_k\rangle. \tag{1.4}$$

The flavour state $|\nu_\alpha\rangle$ defined in Equation 1.1 is created in association with a charged lepton. Provided the CPT operation is invariant, the unitarity of U ensures that the neutrino and lepton associated with a vertex are of the same flavour. This provides the basis of neutrino detection, allowing the experimentalist to infer the neutrino flavour indirectly from the charged lepton found in the final state of charged current quasi elastic (CCQE) interactions.

As a neutrino evolves through space and time, the possibility of observing it in another flavour state becomes non-zero. The propagation of the mass eigenstate must obey the time-dependent Schrödinger equation, the solution of which is a plane wave. To use a plane wave we must make an assumption that the mass eigenstates have equal momentum, however in a more realistic wave packet treatment where this assumption is not needed, the same result is found (See Chapter 8 of [6]). Using the plane wave $\psi = e^{-iE_k t}$ to describe the time evolution, we see that at a time t, the flavour state can be written

$$|\nu_\alpha(t)\rangle = \sum_k U^*_{\alpha k} e^{-iE_k t} |\nu_k\rangle. \tag{1.5}$$

If we now take $|\nu_k\rangle = \sum_k U_{\beta k}|\nu_\beta\rangle$ and insert it into equation 1.5, we find

$$|\nu_\alpha(t)\rangle = \sum_{\beta=e,\mu,\tau} \left(\sum_k U^*_{\alpha k} \psi U_{\beta k} \right) |\nu_\beta\rangle. \tag{1.6}$$

Using the relation $\langle\nu_\alpha|\nu_\beta\rangle = \delta_{\alpha\beta}$, the amplitude for a neutrino born of definite flavour α to be observed as flavour β can be written as

$$A(\nu_\alpha \to \nu_\beta) = \langle\nu_\beta|\nu_\alpha\rangle = \sum_i U^*_{\alpha i} \psi U_{\beta i}, \tag{1.7}$$

where the amplitude for ν_α to be ν_i and ν_i to be ν_β are $U^*_{\alpha i}$ and $U_{\beta i}$ respectively, and ψ is the amplitude for propagation of the neutrino. Now explicitly writing the propagation amplitude ψ, the probability of oscillation becomes

$$P(\nu_\alpha \to \nu_\beta) = |A(\nu_\alpha \to \nu_\beta)|^2 = \sum_{k,j} U^*_{\alpha k} U_{\beta k} . U_{\alpha j} U^*_{\beta j} e^{-i(E_k - E_j)t} \tag{1.8}$$

The energy of the eigenstates enters as a phase term in the probability. Using the assumption of relativistic neutrinos ($p \gg m$ so that $E = |p|$), and that our neutrinos

are created with equal momentum (plane wave approximation), we can write the energy (using a binomial expansion) as

$$E_k = \sqrt{p^2 + m_k^2} = p + \frac{m_k^2}{2p} = E + \frac{m_k^2}{2E}, \tag{1.9}$$

and thus

$$E_k - E_j = \frac{\Delta m_{kj}^2}{2E}, \tag{1.10}$$

where $\Delta m_{kj}^2 \equiv m_k^2 - m_j^2$. Assuming the neutrinos are relativistic so that $t = L$ where L is the propagation length of the neutrino, we can express the phase factor as a function of the difference between the mass states, and the observables E and L

$$P(\nu_\alpha \to \nu_\beta) = |A(\nu_\alpha \to \nu_\beta)|^2 = \sum_{k,j} U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^* e^{-\frac{\Delta m_{kj}^2}{2E} L}. \tag{1.11}$$

1.2.2. Two Flavour Approximation

If the third massive neutrino state is neglected, one can simplify the oscillation formulas and more easily visualise the oscillation phenomena. This is viable because many experiments and scenarios are not sensitive to three neutrino mixing and can be approximated by an effective two neutrino model.

In the case of the simpler two neutrino model, the mixing matrix becomes

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{1.12}$$

which is a simple rotation in 2D space, described by an angle θ. Using 1.12 to compute the probability with Equation 1.11, we arrive at

$$P(\nu_\alpha \to \nu_\beta) = \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right), \tag{1.13}$$

where Δm^2 is the single mass splitting between mass states m_2^2 and m_1^2. Equation 1.14 represents the probability the neutrino will oscillate from flavour α to β after a distance

of L. Using unitarity we can easily express the survival probability as

$$P(\nu_\alpha \to \nu_\alpha) = 1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 L}{4E}\right). \tag{1.14}$$

The parameters L and E are the only ones that the experimentalist can control. Therefore, when planning an experiment, one can select the optimal $\frac{L}{E}$ configuration to maximise the oscillation signal in the detector.

1.2.3. Matter Effects

As a neutrino propagates through vacuum, the formalism described in Section 1.2.1 is sufficient. A neutrino propagating through matter will experience further potentials that affect the oscillation probability, due to the forward scattering of neutrinos on electrons [10].

The neutrino oscillation phenomenon is possible due to the interference between the mass eigenstates as they propagate with different velocities, due to the mass differences. In matter, neutrinos experience an additional potential from interactions with electrons. All three flavours can scatter via the neutral current interaction, however only the electron flavour can interact with the electrons via the charged current mode. As all flavour states feel the same potential from neutral current scattering, a constant potential V_β can be subtracted from the Hamiltonian. This causes a phase difference that enters the Hamiltonian as an extra potential with the form

$$V_\alpha - V_\beta = 2\sqrt{2}G_F E N_e, \tag{1.15}$$

where G_F is the Fermi constant, E is the neutrino energy and N_e is the electron number density of the matter medium. The

Taking solar neutrinos as an example, we can see how this matter induced phase difference can occur in Figure 1.1. Using the two neutrino approximation, the Hamiltonian of the mass eigenstates can be expressed as

$$\mathcal{H} = \mathcal{H}_m + \mathcal{H}_V = \frac{\Delta m^2}{4E}\begin{pmatrix} -\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix} + \begin{pmatrix} V_\alpha - V_\beta & 0 \\ 0 & 0 \end{pmatrix}, \tag{1.16}$$

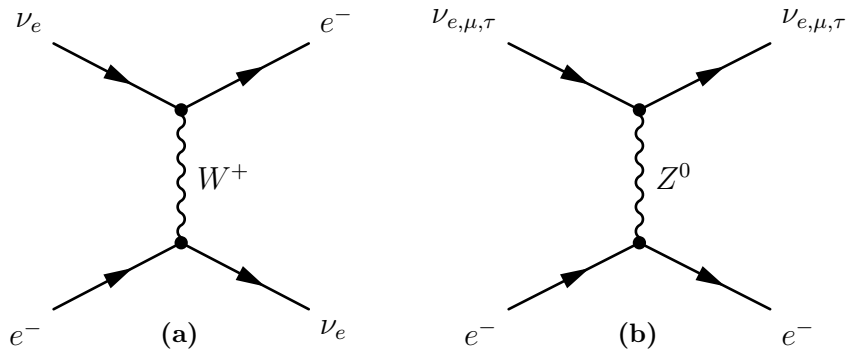


Figure 1.1.: Feynman diagrams for solar neutrino interactions. Neutrinos produced in the solar core may interact via the charged current (a) and neutral current (b) reactions. As there are only electrons in the sun's matter, only the electron neutrino may interact via the CC mode due to the low energy of solar neutrinos. In contrast, all three flavours may couple to the Z^0, allowing NC interactions.

where $\mathcal{H}$ is the total Hamiltonian, $\mathcal{H}_m$ is the vacuum Hamiltonian and $\mathcal{H}_V$ is the potential added to the Hamiltonian due to the presence of matter. Inserting this into the Schrödinger equation and transforming it into the mass eigenstate basis using the mixing matrix, we arrive at

$$i\frac{d}{dt}\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix} = \left[\frac{1}{2E}\begin{pmatrix} m_1^2 + \Delta V\cos^2\theta & \Delta V\cos\theta\sin\theta \\ \Delta V\cos\theta\sin\theta & m_2^2 + \Delta V\sin^2\theta\end{pmatrix}\right]\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix}, \tag{1.17}$$

where $\Delta V = (V_\alpha - V_\beta)$. The mass matrix in 1.17 is no longer diagonal due to the added potential. After diagonalising, it follows that the oscillation probability can be expressed as a function of the matter mass eigenstates and with a matter modified mixing angle

$$P_m(\nu_e \to \nu_\mu) = \sin^2(2\theta_m)\sin^2\left(1.27\Delta m_m^2\frac{L}{E}\right), \tag{1.18}$$

where Δm_m^2 and θ_m are the matter modified mass squared splitting and mixing angle respectively.

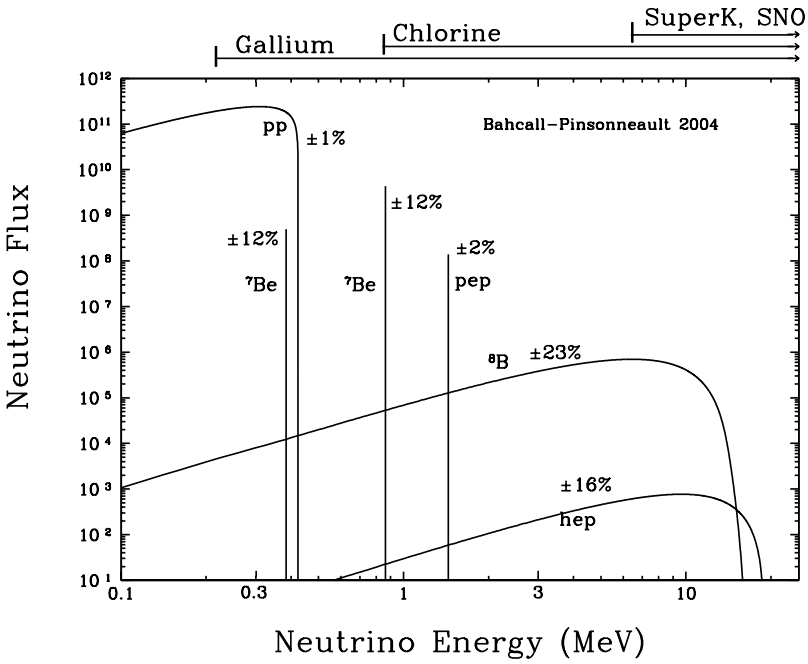


Figure 1.2.: Predicted spectrum of solar neutrinos coming from various processes inside the sun. Image taken from [11].

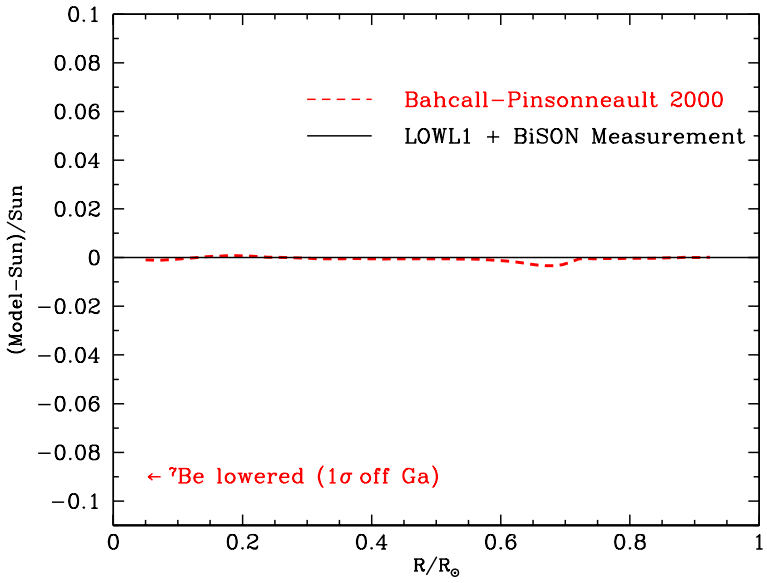


Figure 1.3.: Prediction of solar sound speeds by the SSM compared to experimental measurements. Good agreement is found and thus the missing flux of neutrinos cannot be explained by a flaw in the solar model. Image taken from [12].

1.3. Evidence of Neutrino Oscillation

1.3.1. The Solar Neutrino Problem

In the middle of the 20^{th} century the Standard Solar Model (SSM) was developed to explain the driving mechanisms behind the sun; primarily the proton-proton cycle (see Figure 1.2). A pioneer of the SSM was J. N. Bahcall, who helped shape the model and predict the expected neutrino flux that an experiment could theoretically observe on earth [13]. Working together with R. Davis, their theoretical and experimental efforts established a discrepancy between the predicted and observed rates of neutrinos produced in the sun. As we will see, this discrepancy can be well described by a neutrino oscillation model.

Homestake

The Homestake experiment [14] was the first experiment designed to detect electron-neutrinos from the sun. A large tank containing 390000 l of tetrachloroethylene (C_2Cl_4) was placed 1478 m underground in the Homestake Gold Mine in South Dakota. Running from 1970 to 1994, the experiment measured neutrinos from ^{8}B decays inside the sun using the inverse beta-decay reaction:

$$\nu_e + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + \text{e}^- \tag{1.19}$$

where the chlorine atom inside the detector captures a ν_e creating an unstable isotope of argon. By purging the tank periodically with helium, the argon was removed and the number of neutrino interactions counted by observing the decay of the argon isotope with a proportional counter. The Homestake experiment measured a solar neutrino induced ^{37}Ar production rate of $(2.56 \pm 0.16(\text{stat.}) \pm 0.16(\text{syst.})) \times 10^{-36}\text{s}^{-1}$ per ^{37}Cl atom, which is consistent with a third of the predicted rate from solar models, which predicts $(8.1 \pm 1.2) \times 10^{-36} s^{-1}$ per ^{37}Cl atom. This result provoked criticism over the experiment and the solar model (see Figure 1.3), but both proved to be correct. Thus, the *solar neutrino problem* was born.

Gallium Experiments

There were several experiments designed to measure the solar flux of neutrinos and investigate the deficit seen by the Homestake experiment. These experiments made use of vessels of gallium, due to its low threshold for detection (233 keV), as seen in Figure 1.2. The detection method is via the reaction

$$\nu_e + {}^{71}\mathrm{Ga} \rightarrow {}^{71}\mathrm{Ge} + \mathrm{e}^- \tag{1.20}$$

where the gallium is transmuted into germanium. The germanium is extracted from the vessel and its decay via electron capture is measured using proportional counters.

The first gallium experiments were SAGE [15] and GALLEX [16], which began taking data in the early 90s. GNO [17] was the successor to GALLEX and began taking data in 1998.

SAGE reports, for an analysis of data taken between 1990–2007, a rate of $65.4^{+3.1}_{-3.0}$ (stat.)$^{+2.6}_{-2.8}$(syst.) solar neutrino units (SNU) [18]. GALLEX took data until 1997, and the final result of the analysis measured $77.5 \pm 17.8(\mathrm{stat.}) \pm 6.69(\mathrm{syst.})$ SNU [19]. After running until 2003, GNO measured a rate of $62.9 \pm {}^{5.5}_{5.3}(\mathrm{stat.}) \pm 2.5(\mathrm{syst.})$ SNU.

A weighted combination of all 3 experiments yields 66.1 ± 3.1 SNU, with statistical and systematic errors combined in quadrature. This is approximately 1/2 of the SSM prediction [20].

Kamiokande

The Kamioka Nucleon Decay Experiment [21] started in 1983, and was a 1 kton water Čerenkov detector located 1000 km underground in the Kamioka mine, Japan. After upgrades adding an outer detector in 1986 to help reduce backgrounds, Kamiokande was able to observe ^{8}B neutrinos and measured a flux $0.46 \pm 0.13(\mathrm{stat.}) \pm 0.08(\mathrm{syst.})$ of the SSM prediction. This is in agreement with the measurements obtained by the gallium experiments detailed in Section 1.3.1. As the recoil electron from the neutrino-electron interaction is scattered in the forward direction, Kamiokande was also able to show that the neutrinos they observed were of solar origin.

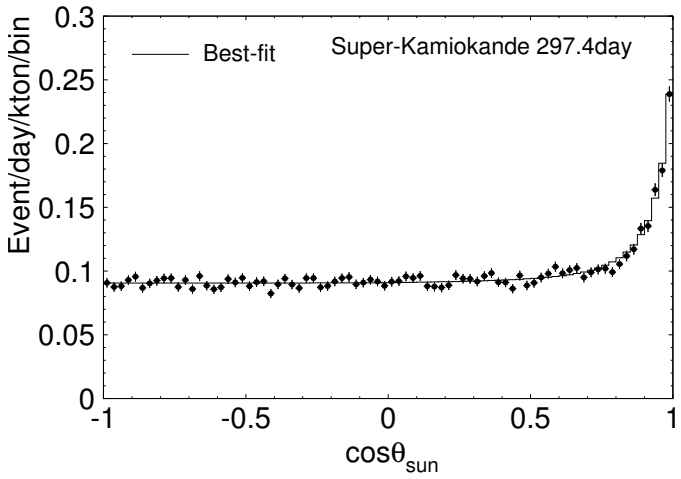


Figure 1.4.: The cosine of the angle between the electron trajectory and the direction of the sun at the time of measurement. A peak at 1 is seen, indicating that the neutrinos detected were coming from the direction of the sun. The peak is broadened due to the multiple scattering of the electron. Figure taken from [22].

Super-Kamiokande

The Super-Kamiokande (SK) water Čerenkov detector (see Section 2.5 for more detail) gave strong evidence that the deficit of neutrinos from the sun was indeed true [22].

By looking for ^{8}B decay neutrinos by observing recoil electrons of energy greater than 6.5 MeV, the SK detector measured a total flux of $(2.42 \pm 0.06(\text{stat.})^{+0.01}_{-0.07}(\text{syst.}))$ $\times 10^6 \text{cm}^{-2}\text{s}^{-1}$. This result is 36 % of the expected neutrino flux predicted by the solar model. SK was able to confirm that the neutrinos seen inside the fiducial volume of the detector were indeed coming from the sun due to the direction of the electron's Čerenkov cone, as seen in Figure 1.4.

SNO

The Sudbury Neutrino Observatory (SNO) was a water Čerenkov detector that studied the flux of high energy neutrinos coming from ^{8}B decay in the sun [4].

SNO was a spherical acrylic vessel of 6 m radius containing 1000 t of heavy water. Located 2 km underground in the Creighton Mine in Sudbury, Canada, it operated from 1999 to 2006 and was able to measure the total flux of neutrinos coming from the sun, due to the extra interactions that are possible with heavy water.

The three main processes by which SNO measured the ^{8}B decay neutrino flux are:

$$\nu + \mathrm{d} \rightarrow \mathrm{e}^- + \mathrm{p} + \mathrm{p} \text{ (charged current)} \tag{1.21}$$

$$\nu + \mathrm{d} \rightarrow \nu + \mathrm{p} + \mathrm{n} \text{ (neutral current)} \tag{1.22}$$

$$\nu + \mathrm{e}^- \rightarrow \nu + \mathrm{e}^- \text{ (charged \& neutral current)} \tag{1.23}$$

where process 1.21 is charged current deuteron breakup, which can only happen with a ν_e as solar neutrinos do not have the required energy to produce a μ or τ particle. Therefore, this reaction is sensitive only to the flux of ν_e. Process 1.22 is neutral current deuteron breakup, where the Z^0 can couple to all three active neutrino flavours equally. Process 1.23 is the elastic scattering on an electron. This also can happen with any active neutrino flavour, but the cross-section for ν_e scattering is larger than that for ν_μ or ν_τ by a factor of approximately 6.5.

Using these three reactions, SNO was sensitive to the total flux of ^{8}B decay neutrinos. As only ν_e neutrinos are produced in the sun, a flux of ν_μ or ν_τ detected would be strong evidence that ν_e produced in the sun were changing flavour. Indeed, using processes 1.21 and 1.22, a flux of non-electron flavour neutrinos were found.

Assuming no oscillation, the ratio of fluxes measured by using the processes 1.21 and 1.22 should be $\frac{\Phi(\nu_e)}{\Phi(\nu_e)+\Phi(\nu_\mu,\nu_\tau)} = 1$. In fact, SNO measured [23]

$$\frac{\Phi(\nu_e)}{\Phi(\nu_e) + \Phi(\nu_\mu, \nu_\tau)} = 0.340 \pm 0.023(\text{stat.})^{+0.029}_{-0.031}(\text{syst.}), \tag{1.24}$$

where there is clearly a flux of ν_μ and/or ν_τ. This provided solid evidence that the ν_es born in the sun were changing flavour.

As the oscillation phenomena changes only the flavour of the neutrino, the total flux $\Phi(e, \mu, \tau)$ should remain unaffected. SNO was sensitive to the total flux, and measured $(4.94 \pm 0.21\ (\text{stat.})^{+0.38}_{-0.34}(\text{syst.})) \times 10^6 \text{cm}^{-2}\text{s}^{-1}$, which is in good agreement with the prediction of the standard solar model: $5.49^{+0.95}_{-0.81} \times 10^6 \text{cm}^{-2}\text{s}^{-1}$.

KamLAND

The Kamioka Liquid Scintillator Antineutrino Detector (KamLAND) [24] is a detector designed to be sensitive to the solar neutrino oscillation parameters θ_{12} and Δm^2_{12} using anti-neutrinos from nuclear reactors. Located underground near Toyama, Japan, it sits an average distance of 180 km from multiple reactor sources. The reactors produce a flux of $\bar{\nu_e}$, which are detected inside KamLAND using the inverse beta-decay reaction:

$$\bar{\nu_e} + \mathrm{p} \rightarrow \mathrm{e}^+ + \mathrm{n} \tag{1.25}$$

which produces a distinct prompt and delayed coincidence signal used to select $\bar{\nu_e}$. The positron will annihilate quickly and produce a prompt signal, followed by the neutron thermalising and then being captured by a hydrogen nucleus, producing a deuteron and a distinct 2.2 MeV photon, which constitutes the delayed signal. This signature coincidence is the principal detection method for KamLAND, which was first used in the discovery of the neutrino [25].

In 2008, the KamLAND collaboration published their precision measurement of the solar mixing parameters [26]. The survival probability in Figure 1.5 shows a clear sinusoidal distribution consistent with neutrino oscillation. Figure 1.6 shows the allowed regions of said parameters when overlayed with the results of solar experiments. To date, KamLAND and the solar experiments have provided the most sensitive measurement of θ_{12} and Δm^2_{12}. KamLAND in particular has acted to confirm the solar neutrino oscillation phenomenon using purely a terrestrial experiment.

1.3.2. The Atmospheric Neutrino Anomaly

Atmospheric neutrinos are produced in the upper atmosphere by cosmic rays (See Chapter 11 of [6]). This occurs when a cosmic ray (typically a proton) collides with an air molecule, causing a shower of particles with an energy peak on the order of $\sim$1GeV. The pions produced by cosmic rays decay mainly into muons and muon neutrinos, as show in Figure 1.7

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \pi^- \rightarrow \mu^- + \bar{\nu}_\mu. \tag{1.26}$$

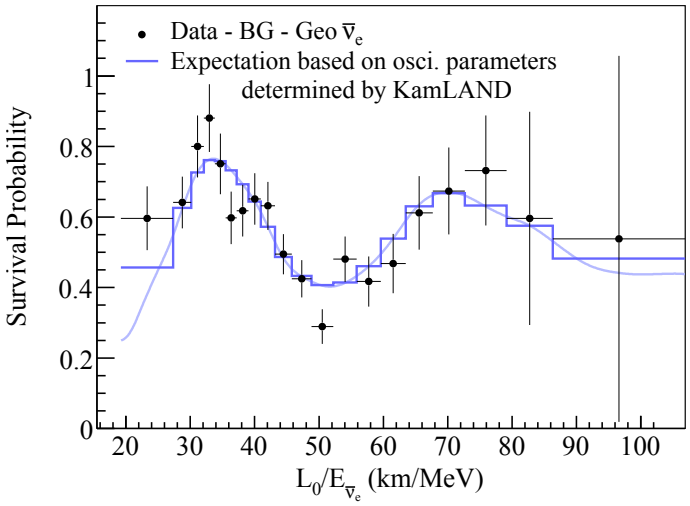


Figure 1.5.: Survival probability $P(\bar{\nu}_e \to \bar{\nu}_e)$ as measured by KamLAND. The data favours the neutrino oscillation model and disfavours alternative models such as neutrino decay and decoherence. Figure taken from [26].

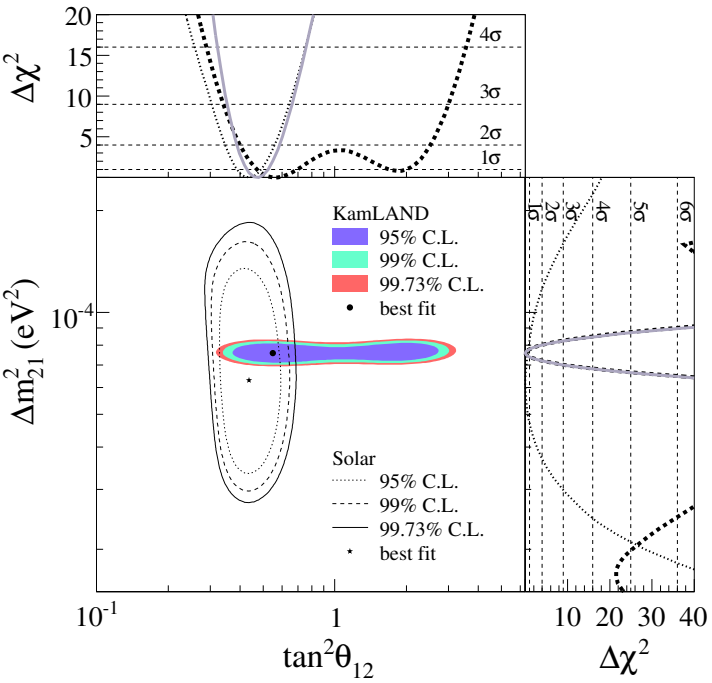


Figure 1.6.: Allowed regions of $\tan^2\theta_{12}$ and Δm^2_{12} constrained by KamLAND reactor data and the combined allowed regions of solar neutrino experiments. Figure taken from [26].

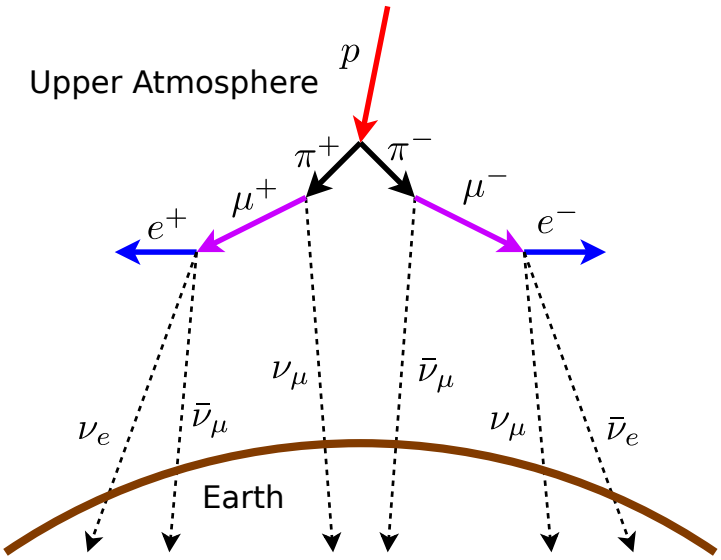


Figure 1.7.: Diagram showing the process by which neutrinos are produced by cosmic rays. A cosmic ray will enter the upper atmosphere and interact with a molecule of air. This causes a shower of pions, which go on to decay into $\mu^\pm$. These muons travel further towards the earth and eventually decay into $e^\pm$. At every stage of this process, various flavours of neutrinos and anti-neutrinos are produced.

For energies that are below 1 GeV, the muons will mostly decay into electron neutrinos and electrons

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu, \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu. \tag{1.27}$$

It can be seen from reaction 1.26 and 1.27 that the flux ratios of atmospheric neutrino flavours is different for energies above and below 1 GeV.

Super-Kamiokande

In 1998 the Super-Kamiokande collaboration published their results of an analysis of the angular distributions of atmospheric muon- and electron-like neutrino events [5]. Figure 1.8 shows the atmospheric neutrino data plotted with the MC prediction from atmospheric models and also the MC prediction for oscillations. The muon-like events are consistent with neutrino oscillation, whereas the electron-like events are consistent with no oscillation.

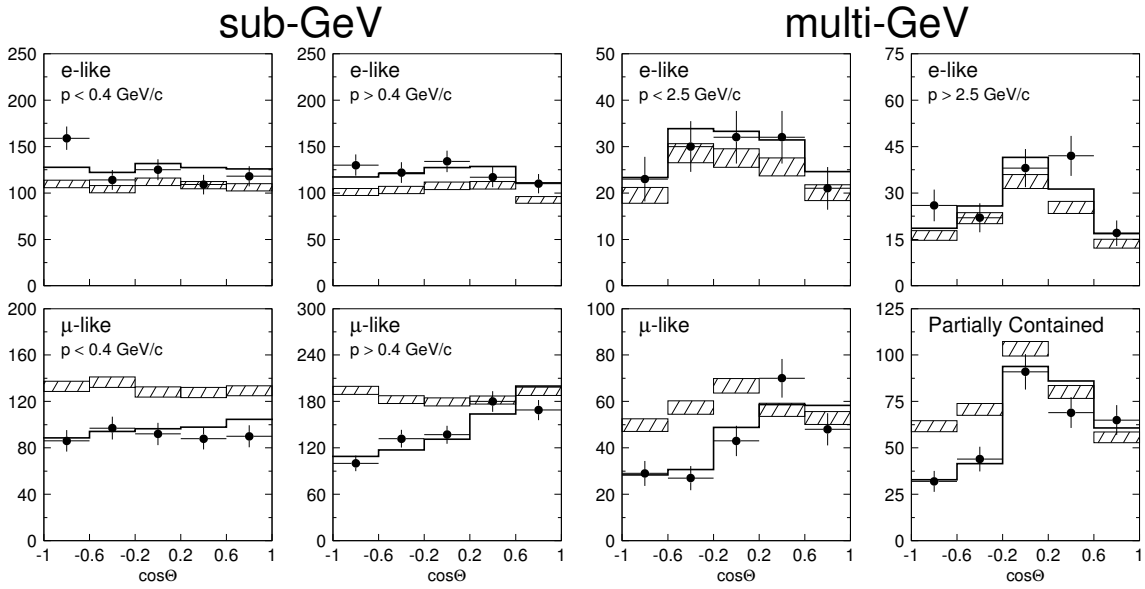


Figure 1.8.: SK atmospheric neutrino analysis. Data is divided into $E < 1$ GeV and $E > 1$ GeV to show the energy dependence of the atmospheric neutrino flux ratios as discussed in Section 1.3.2. Black points are data, solid black lines are the MC prediction for oscillations, and the hatched areas are the no oscillation prediction. Figure taken from [5].

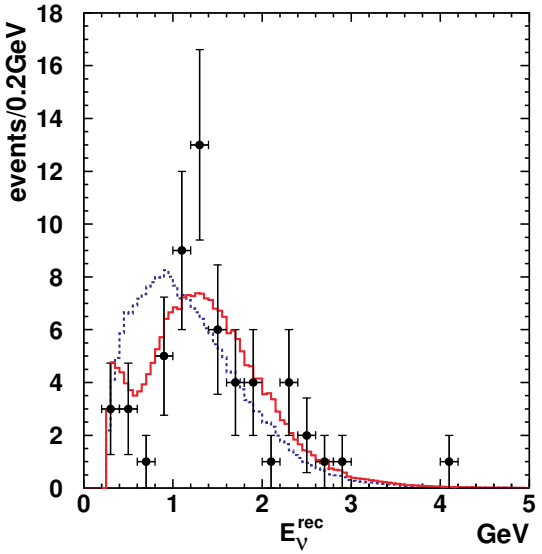


Figure 1.9.: Neutrino energy reconstructed from single-ring, μ-like events. The black points with error bars are the data, the dashed line is the MC prediction for the no oscillation hypothesis, and the solid line is the best fit spectra assuming the neutrino oscillation model. Figure taken from [27].

K2K

The KEK-to-Kamioka experiment (K2K) was a long-baseline neutrino experiment that was the first experiment to use an accelerator-made beam of ν_μ neutrinos to study and confirm the oscillation results seen by atmospheric experiments [27]. Using a baseline of 250 km, the experiment propagated the neutrino beam to the SK detector, where after a 5 year data taking period, 112 beam ν_μ neutrino events were detected compared to an MC expectation of $158^{+9.2}_{-8.6}$. This corresponds to the no oscillation hypothesis being excluded to 4.3 σ, which, along with the distortion of the energy spectrum seen in Figure 1.9, was compelling evidence and terrestrial confirmation of the atmospheric neutrino experimental results.

1.3.3. Entering the Precision Era

With the establishment of the neutrino oscillation phenomena, experiments were designed to measure the parameters describing the mixing with precision.

CHOOZ

The CHOOZ experiment [28] was a long baseline reactor neutrino experiment located in Chooz, France. The experiment studied a nearly pure $\bar{\nu}_e$ source 1 km away from the pressurized water reactors that produced them.

The detector was located underground with a 300 mwe (metre-water-equivalent) overburden, and comprised of a 5.5 m tall, 5.5 m diameter 5 t vessel of liquid scintillator loaded with gadolinium. The gadolinium allowed the detector to tag the delayed neutron-capture produced in an inverse beta decay reaction.

CHOOZ measured a ratio of measured to expected $\bar{\nu}_e$ flux of:

$$R = 1.01 \pm 0.028 \pm 0.027, \tag{1.28}$$

which excludes the $\nu_\mu \leftrightarrows \nu_e$ oscillation hypothesis as an explanation of the atmospheric neutrino anomaly. As shown in Figure 1.10, CHOOZ provided an upper limit on θ_{13}.

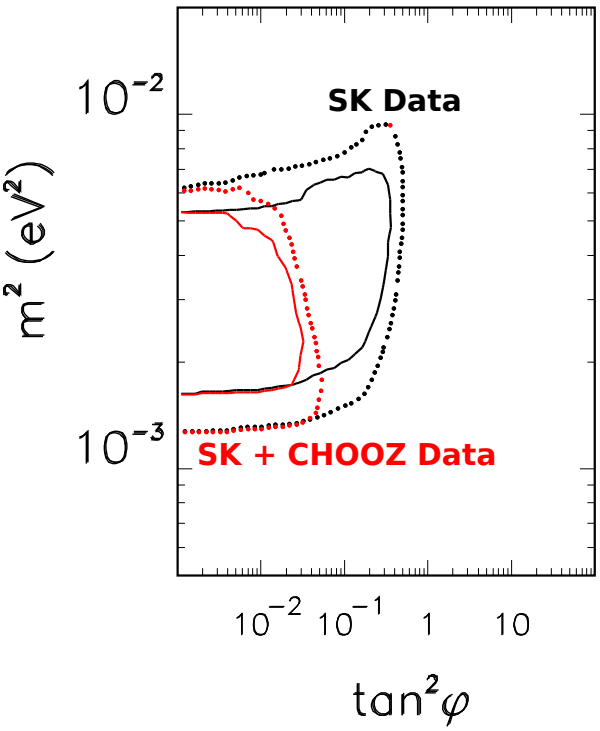


Figure 1.10.: CHOOZ limit on ν_e mixing, where $\phi \equiv \theta_{13}$. Figure adapted from [28].

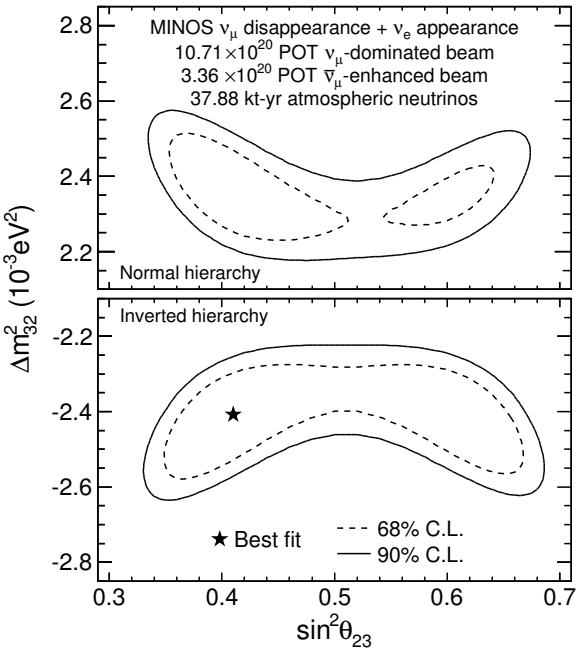


Figure 1.11.: MINOS analysis of atmospheric and accelerator neutrino data. The analysis finds the best fit value in the inverted hierarchy. Figure taken from [29]

MINOS

MINOS is a long baseline experiment using the NuMI beamline and a near detector at Fermilab, Chicago and with a far detector 735 km away in Minnesota [30].

The near and far detectors are both scintillator based with steel plating between layers. The near detector has a mass of 980 t and is 100 m underground, whereas the far detector has a mass of 5.4 kt and is 716 m underground. The detectors are magnetised to allow discrimination between neutrinos and anti-neutrinos.

A recent analysis of the MINOS data set from 2006 to 2012, comprising of accelerator neutrinos and anti-neutrinos, along with atmospheric neutrino data [29] yields

$$|\Delta m^2_{32}| = [2.28 - 2.46] \times 10^{-3}\text{eV}^2 \text{ (68\% C.L.) and}$$
$$\sin^2\theta_{23} = 0.35 - 0.65 \text{ (90\% C.L) for normal hierarchy}$$

and

$$|\Delta m^2_{32}| = [2.32 - 2.53] \times 10^{-3}\text{eV}^2 \text{ (68\% C.L.) and}$$
$$\sin^2\theta_{23} = 0.34 - 0.67 \text{ (90\% C.L) for inverted hierarchy.}$$

The contours constructed for both mass hierarchies can be seen in Figure 1.11, where the best fit value is in the inverted hierarchy.

Daya Bay

Daya Bay is a neutrino oscillation experiment in China, that observes the oscillations of anti-neutrinos produced from surrounding nuclear reactors [31]. The Daya Bay experiment is comprised of 6 liquid scintillator detectors doped with gadolinium, which are arranged into two near and one far hall sites. The detectors are equipped with PMTs to detect the inverse beta decay reaction produced from anti-neutrinos; the same method as KamLAND (Section 1.3.1).

Between December 2011 and February 2012, Daya Bay collected 10416 reactor anti-neutrinos in the far hall. By comparison with the near hall prediction, there was a 6% deficit compared to the expected number of events. An analysis based on the neutrino rate (see Figure 1.12) measures $\sin^2 2\theta_{13} = 0.092 \pm 0.016(\text{stat.}) \pm 0.005(\text{syst.})$, and rejects the no-oscillation hypothesis at a significance of $5.2\,\sigma$. This was the first precision measurement of θ_{13}.

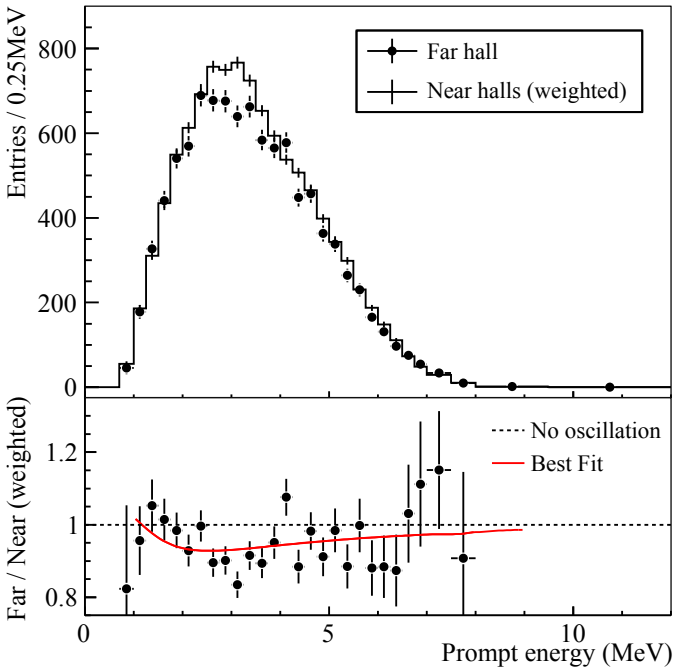


Figure 1.12.: Daya Bay prompt energy spectrum. Figure taken from [31].

1.4. Current Status of Neutrino Oscillations

Neutrino oscillation physics has entered the precision measurement era. The non-zero determination of θ_{13} by T2K [8] and reactor experiments like Daya Bay [9] has opened up the possibility of observing CP violation in the lepton sector, which requires yet more precise knowledge of all the oscillation parameters. Although we have precise measurements of the atmospheric parameters by MINOS [30] and T2K [32], further sensitivity is needed, along with the determination of the neutrino mass hierarchy, as shown in Figure 1.13. Matter effects in the sun have given us the sign of the solar mass splitting Δm^2_{12}, however only the magnitude of Δm^2_{32} is known. It is expected that upcoming experiments such as PINGU [33], with longer baselines than current generation experiments, will determine the mass hierarchy. The current measurements of the oscillation parameters determined by a global fit to oscillation data is presented in Table 1.1.

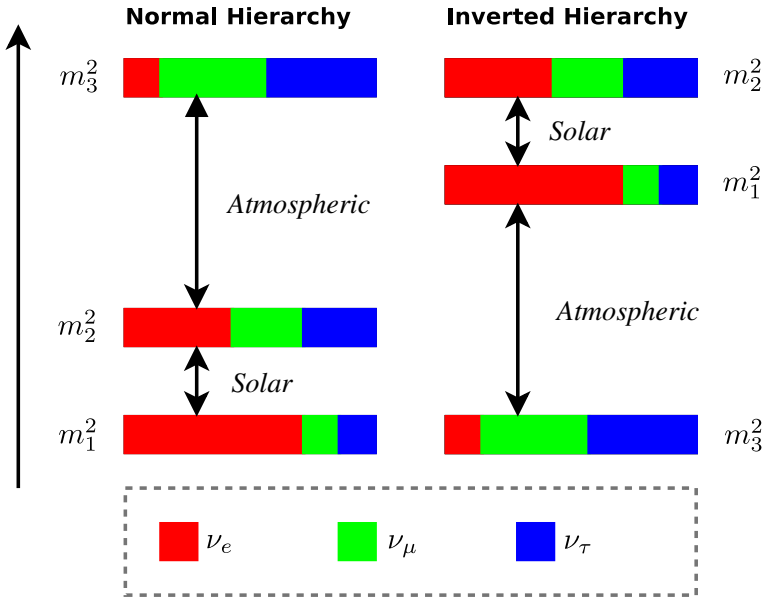


Figure 1.13.: Diagram showing the two different configurations of neutrino mass eigenstates that constitute the neutrino mass hierarchy. On the left is the normal hierarchy, where the third mass state m_3^2 is the heaviest. In contrast, on the right is the inverted hierarchy, where m_3^2 is the lightest mass state of the three.

Parameter	Best Fit	1σ range
$\delta m^2/10^{-5}\text{eV}^2$ (NH or IH)	7.54	7.32 - 7.80
$\sin^2\theta_{12}/10^{-1}$ (NH or IH)	3.08	2.91 - 3.25
$\Delta m^2/10^{-3}\text{eV}^2$ (NH)	2.44	2.38 - 2.52
$\Delta m^2/10^{-3}\text{eV}^2$ (IH)	2.40	2.33 - 2.47
$\sin^2\theta_{13}/10^{-2}$ (NH)	2.34	2.16 - 2.56
$\sin^2\theta_{13}/10^{-2}$ (IH)	2.39	2.18 - 2.60
$\sin^2\theta_{23}/10^{-1}$ (NH)	4.25	3.98 - 4.54
$\sin^2\theta_{23}/10^{-1}$ (NH)	4.37	4.08 - 4.96 $\oplus$ 5.31 - 6.10

Table 1.1.: Global fit of neutrino oscillation data. Δm^2 is defined as $m_3^2 - (m_1^2 + m_2^2)/2$. Table contents taken from [34].

Chapter 2.

The Tokai-to-Kamioka Experiment

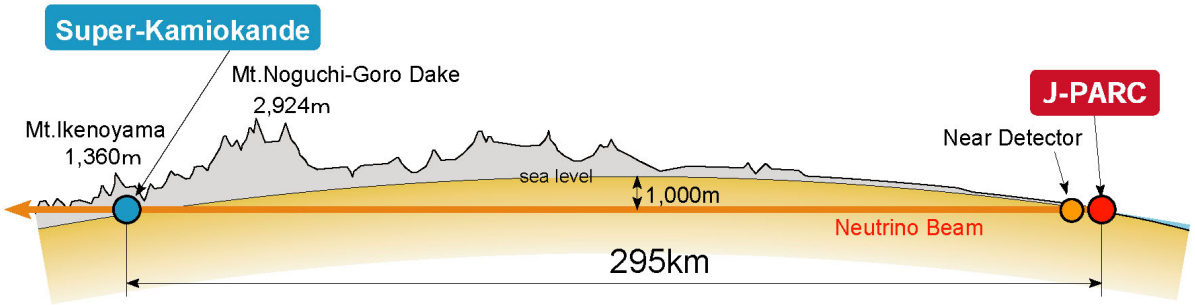


Figure 2.1.: Diagram of T2K experimental setup, showing J-PARC, where the beam is produced and measured by the near detector, and also the far detector Super-Kamiokande. Image taken from [35].

The Tokai-to-Kamioka (T2K) experiment is a second generation long-baseline neutrino oscillation experiment. The experiment consists of an intense beam of ν_μ created at J-PARC, Ibaraki prefecture, Japan, which propagates 295 km to the Super-Kamiokande (SK) detector in Gifu prefecture. The experimental axis is configured 2.5 degrees off-axis from the beam, which creates an neutrino energy spectrum peaked at 0.6 GeV with a reduced high energy component to suppress backgrounds as shown in Figure 2.2. T2K has sensitivity to the θ_{23}, θ_{13}, Δm^2_{23} and δ_{cp} oscillation parameters through the ν_μ disappearance and ν_e appearance channels measured at the far detector. Neutrino beam flux is simulated by Monte Carlo and constraints from external data, however a near detector (ND280) located 280 m from the neutrino production point has been constructed to measure the unoscillated beam content, which further constrains the beam

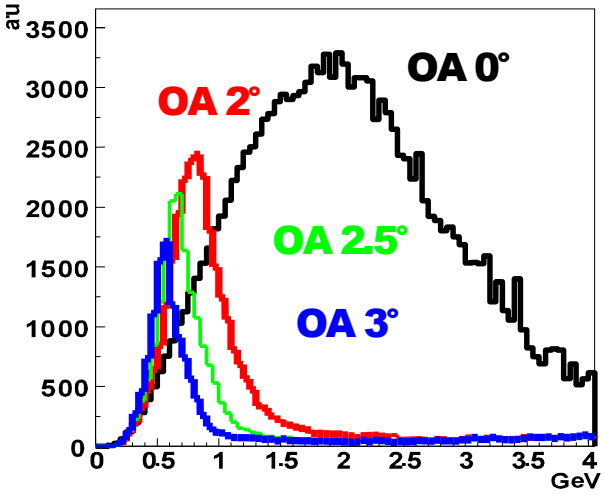


Figure 2.2.: The neutrino energy spectra simulated at varying off-axis angles. As the off-axis angle increases, the high energy neutrinos become suppressed, forming a narrow energy peak. Figure taken from [36].

flux prediction and also correlated cross-section parameters between the near and far detectors.

2.1. Neutrino Interactions

In T2K, the charged current quasi-elastic (CCQE) interaction channel is used for the appearance and disappearance oscillation analyses. As seen in Figure 2.5, at the T2K beam peak energy, the main charged current interaction mode is quasi-elastic scattering (QE), with resonant production (RES) interactions contributing at higher energies, and finally in the high energy tail of beam some deep inelastic scattering (DIS) interactions occur. These modes are categorized by the configuration of charged particles in the final state. Feynman diagrams showing CCQE and RES interactions can be seen in Figure 2.3.

The CCQE interaction mode involves a W boson being absorbed by a neutron, resulting in a single charged lepton and, depending on the kinematics, a recoil proton:

$$\nu_\alpha + \mathrm{n} \rightarrow l_\alpha + \mathrm{p}. \tag{2.1}$$

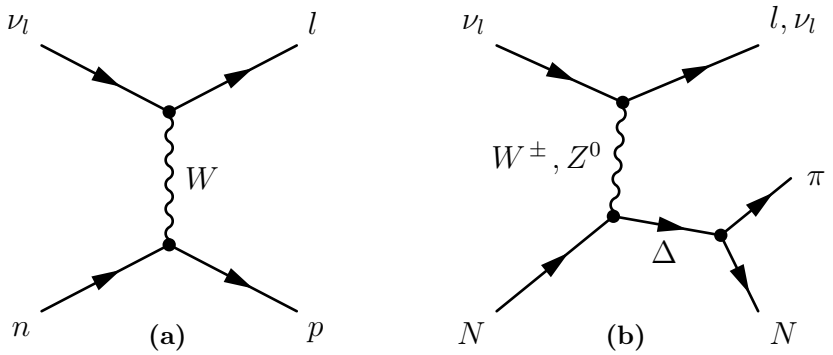


Figure 2.3.: Feynman diagrams for (a) charged current quasi elastic scattering and (b) charged current / neutral current resonance interactions.

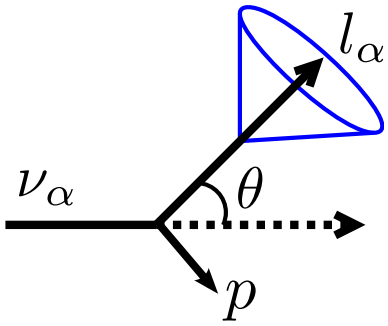


Figure 2.4.: A neutrino interacts with nucleon, producing a charged lepton whose momentum angle to the neutrino trajectory are dependent on the initial neutrino energy. In the case of SK, the charged lepton produces a ring of Čerenkov radiation. A recoil proton may also be visible.

Assuming that the neutron is at rest, the reconstructed neutrino energy can be expressed in terms of the lepton kinematics alone [37]:

$$E_{rec} = \frac{m_p^2 - m_n^2 - m_l^2 + 2E_l m_n}{2(m_n - E_l + |\vec{p_l}| \cos\theta_l)} \tag{2.2}$$

where m_n, m_p and m_l are the neutron, proton and lepton masses respectively, $E_l = \sqrt{p_l^2 + m_l^2}$ is the lepton energy, p_l the lepton momentum and θ_l the angle between the lepton and neutrino beam as shown in Figure 2.4. It is the fact that the neutrino energy can be inferred from the kinematics of the lepton that makes CCQE the preferred interaction mode to study oscillations using both SK and ND280.

Other neutrino interaction modes create a background for the oscillation signal due to events being mistaken for CCQE in event reconstruction. Dominant modes that produce

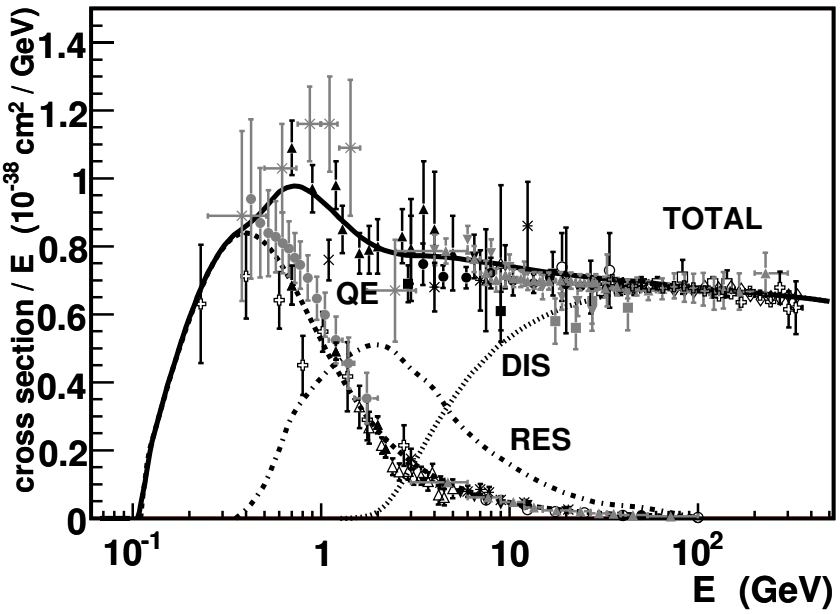


Figure 2.5.: Neutrino cross-section as a function of neutrino energy. The total cross-section is subdivided into the quasi-elastic, resonant and deep inelastic scattering interaction modes. The T2K neutrino beam peaks at 0.6 GeV. Figure taken from [39]

background are the resonant and coherent pion production channels, where the neutrino interacts with the nucleus, resulting in pion production. These pions can cause the event to be misidentified as CCQE if the pion is mistaken for a muon, or if it is not observed. Another channel that causes background is the DIS mode, where the neutrino interacts with the partons within the nucleus, causing the nucleus to break apart and produce many particles in its final state. A large background producing channel for ν_e appearance is the neutral current π^0 (NCπ^0) mode, where a π^0 is produced and can only be detected when it decays into 2 photons. At SK, these 2 photons may overlap, or one may be misreconstructed, causing the event to look like an electron-like CCQE event.

The final state of the neutrino interaction categorizes the event mode, however there are additional complications of multi-nucleon effects and final state interactions [38] which alter the observed final state of the interaction, and so motivate the inclusion of systematics to account for the uncertainty in the true nature of the event.

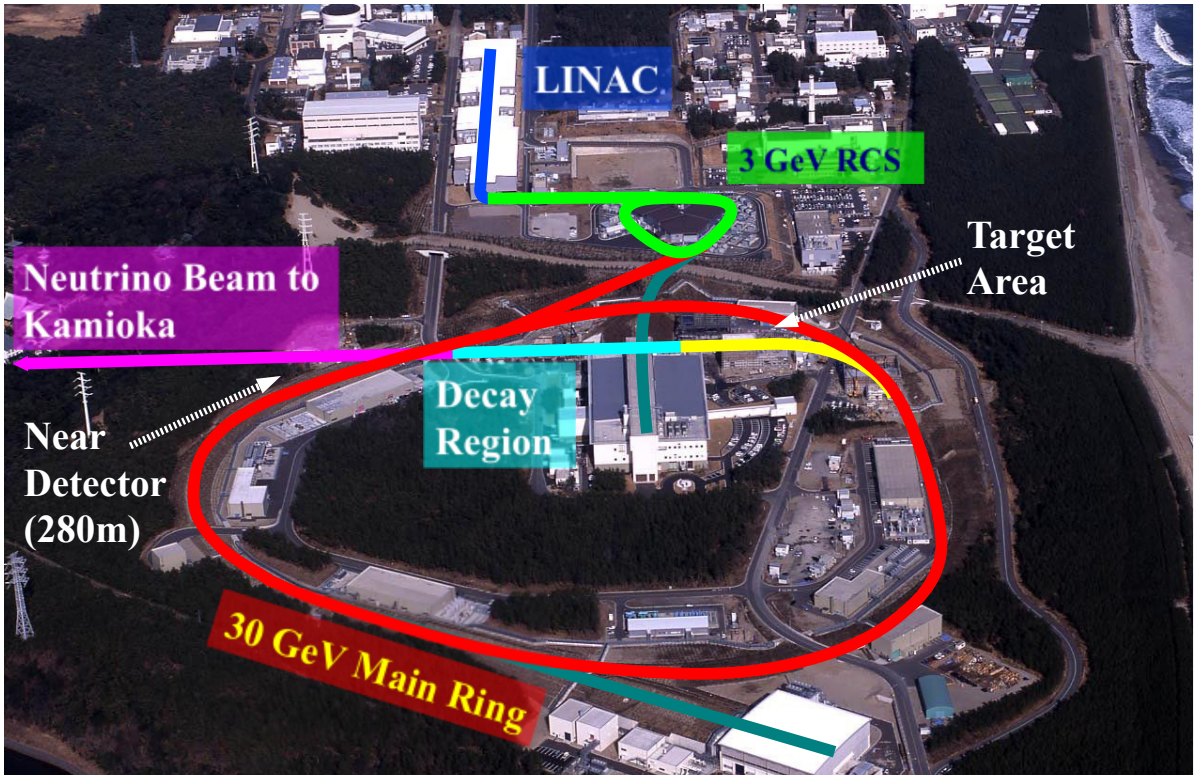
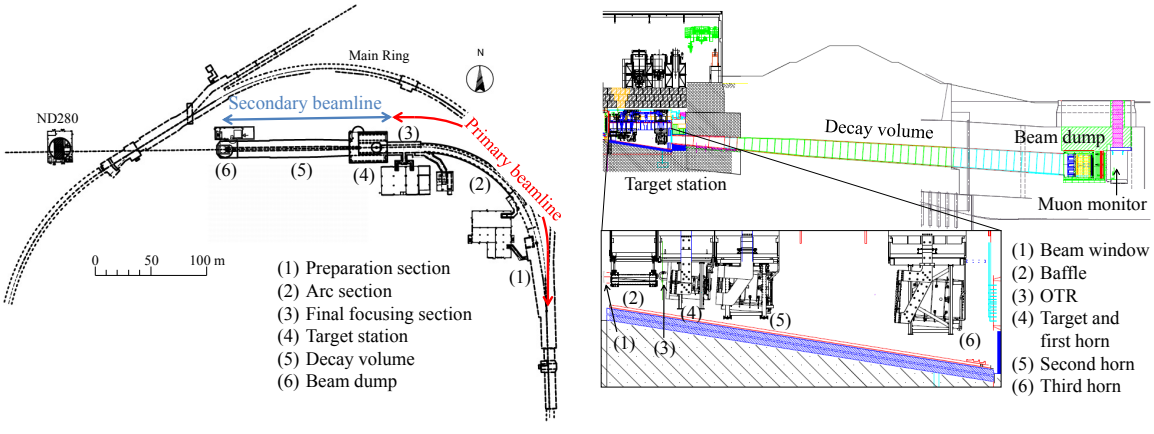


Figure 2.6.: J-PARC aerial photograph, with all the elements relevant to neutrino beam production highlighted. Figure taken from [40].



(a) Diagram showing the neutrino beamline components responsible for extracting protons from the MR and feeding them to the proton target station.

(b) Diagram showing the proton target station, decay volume and beam dump/muon monitor.

Figure 2.7.: Diagrams showing the beamline components relevant to neutrino production. Figures taken from [35].

2.2. Beamline

The J-PARC facility consists of three accelerators that produce an intense bunched proton beam. The three stages of accelerators are a linear accelerator (LINAC), a rapid-cycling proton synchrotron (RSC) and a main ring synchrotron (MR) [35]. These sections are illustrated in Figure 2.6. An H^- beam is accelerated to 180 MeV with the LINAC, and is converted into an H^+ beam by charge-stripping foils. The H^+ beam is then injected into the RCS where they are accelerated to 3 GeV at 25 Hz. For the final stage of acceleration, the protons enter the MR where they reach 30 GeV and have an eight bunch time structure.

To produce a neutrino beam, the eight proton bunches are deflected out of the MR by superconducting magnets and are then extracted by five kicker magnets, as shown in Figure 2.7a. The bunches are separated by 500 ns, and this information is communicated to the far detector via a GPS system in order to trigger on the bunches.

The bunches are then incident upon a proton target which is comprised of a graphite rod of length 91.4 cm and diameter 2.6 cm which is cooled by a helium gas flow. The protons interact with the graphite to produce large quantities of mesons, which are mostly pions. The mesons are focussed into a beam by three magnetic horns [41] operating at 250 kA, and are directed into a 96 m decay volume filled with helium, as shown in Figure 2.7b. The magnetic horns consist of two inner and outer conductors that form a closed volume, which contains toroidal magnetic field. Charged particles are deflected this field, which varies as 1/distance from the horn axis. The current driving the magnetic horns can be changed to deflect positive and negatively charged mesons, to create an anti-neutrino or neutrino beam respectively. Assuming we are making a neutrino beam, as the mesons travel through the decay volume, they decay to produce muon neutrinos via

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.3}$$

and

$$K^+ \rightarrow \mu^+ + \nu_\mu. \tag{2.4}$$

However, the resulting neutrino beam is not pure and there is a small contamination of ν_e due to

$$K^+ \rightarrow \pi^0 + e^+ + \nu_e \tag{2.5}$$

and

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu. \tag{2.6}$$

These neutrinos are now the T2K neutrino beam with a peak energy of around 0.6 GeV. The muons produced in the meson decays are stopped in a beam dump at the end of the decay volume, and are also measured by muon monitors to allow for indirect measurements of the neutrino beam position and flux.

2.3. Off-Axis Near Detector (ND280)

The near detector (ND280) is a magnetised tracking detector composed of several independent sub-detectors (see Figure 2.8). The subsystems provide a fiducial interaction volume, tracking, calorimetry, background measurement and cosmic veto capabilities. The entire detector sits within a homogeneous 0.2 T magnetic field produced by a refurbished magnet donated from CERN.

ND280 sits 280 m from the proton target, 2.5° off-axis with respect to the beam. The motivation to build the near detector was to provide constraints on beam flux parameters, and make valuable neutrino interaction cross-section measurements on oxygen, lead and carbon.

ND280 comprises of two main sections: a π^0 detector and tracking detector, optimised to identify π^0 particles and precise measurements of charged particles respectively. The tracker contains three time projection chambers and two fine grained detectors, which sit within a basket. The basket sits inside the magnet, and is surrounded by electromagnetic calorimeters, which are attached to the inner walls of the magnet.

Also at the near detector complex, but in the on-axis component of the beam, is the interactive neutrino grid detector (INGRID). This detector is designed to monitor the on-axis beam component and contribute to the understanding of the neutrino beam flux and stability.

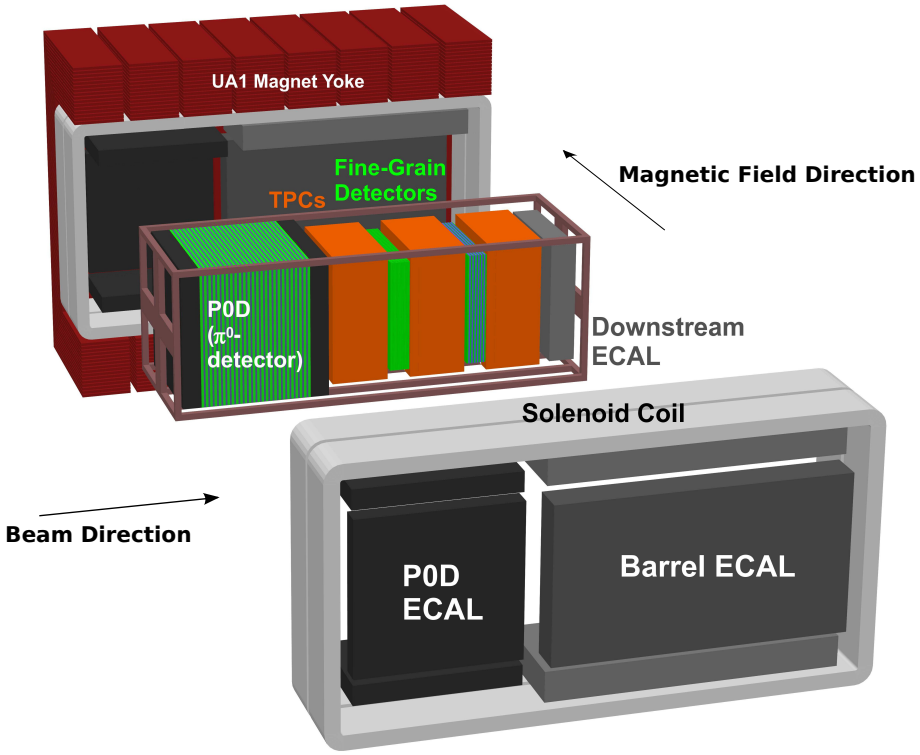


Figure 2.8.: An exploded view of the ND280 near detector. The inner section sits inside a basket, and contains the FGD and TPC tracking subdetectors, π^0 detector and the downstream ECal. The UA1 magnet can be opened and closed around the basket, and is lined with the barrel and P0D ECal modules. Not shown is the side muon range detector (SMRD), which is inter-spaced between the iron pieces of the magnet yoke. Figure taken from [35].

By far the most common component in the ND280 and INGRID detectors is the Multi-Pixel Photon Counter (MPPC), which are used to read out scintillation light from plastic scintillator bars into the data acquisition system. MPPCs are used in all components besides the TPC. Therefore, it is logical to introduce MPPCs first.

2.3.1. Multi-Pixel Photon Counters

Multi-Pixel Photon Counters were selected for use in the T2K near detector complex because they are both small and able to operate in a magnetic field, unlike photomultiplier tubes (PMT). An example of a single MPPC can be seen in Figure 2.9. They are composed of a grid of avalanche photodiodes (APD) [42], with an active area of 1.3 mm^2 and a pixel size of 50 μm^2. An APD consists of a p-n junction that is reverse biased to create

an electric field at the junction. Each of the 667 pixels operates independently in Geiger mode, where the reverse bias voltage is above the breakdown voltage of the APD. When a photon is absorbed, it creates an electron-hole pair that is drifted inside the electric field. Gieger mode ensures that the electron will cause a cascade of electron-hole pairs, which amplifies the signal. The total photodiode signal obtained is a summed contribution from all pixels that fire when a photon is incident on the MPPC. An example of a typical signal from an MPPC can be seen in later on in Figure 3.5. The charge Q_{pixel} obtained from a single pixel can be calculated by:

$$Q_{pixel} = C_{pixel}(V_{op} - V_{bd}) \tag{2.7}$$

where C_{pixel} is the capacitance of the pixel, V_{op} is the operating voltage and V_{bd} is the breakdown voltage of the pixel, which is typically 1-2 V below V_{op}. With an operating voltage of 70 V, and a pixel capacitance 90 fF the photoelectron gain per pixel is of the order of 1×10^6. The MPPC gain depends on both V_{op} and T

$$G = \frac{C_{pixel}}{e[V_{op} - (V_{bd}T_0 + (T - T_0)\frac{dV_{bd}}{dT})]} \tag{2.8}$$

Each MPPC couples to a wavelength shifting (WLS) fibre via a plastic ferrule. The WLS fibre serves as a wave-guide and are inserted into the scintillator bars that make up the active volumes of the near detector complex. As light from the scintillator enters the WLS fibre, it is absorbed by the WLS molecule and re-emitted isotropically. This mechanism helps contain photons that would otherwise travel straight through the fibre.

A total of 64,000 MPPCs were manufactured for the T2K experiment [43]. An extensive study of the performance of the MPPCs used in the T2K experiment can be found in [44].

MPPCs undergo crosstalk between pixels and after-pulsing effects [45]. Crosstalk occurs when a photon produced by an avalanche in one pixel causes an adjacent pixel to trigger. This is troublesome because the number of pixels triggered is intended to be proportional to the intensity of light, and so crosstalk can alter this expectation.

After-pulsing is when charge becomes trapped inside the semiconductor as the pixel fires, which later becomes untrapped, causing a delayed discharge. These crosstalk and after-pulsing effects must be accounted for in the calibration.

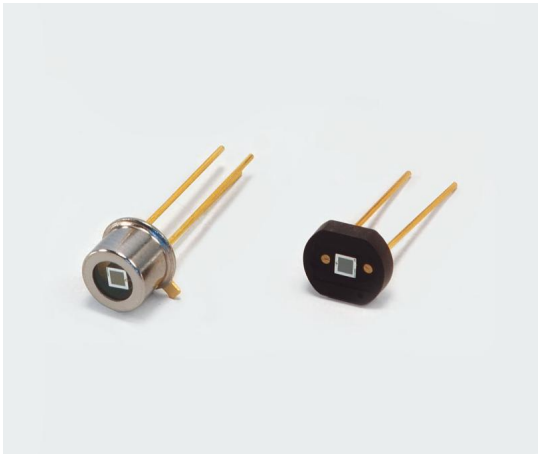


Figure 2.9.: Image of MPPC devices. Image taken from [46].

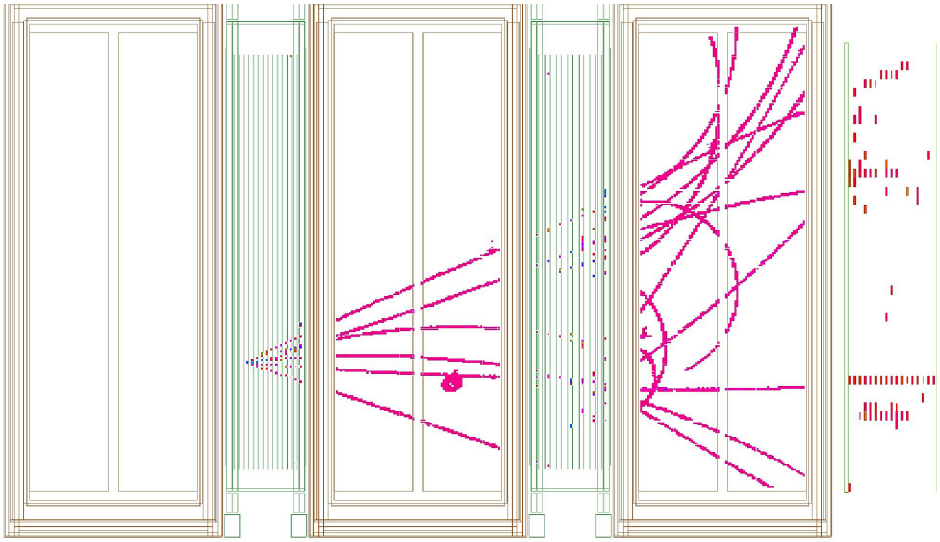


Figure 2.10.: Event display in the yz plane of a neutrino interaction inside the FGD that traverses the full tracker. Visible are the three TPCs, two FGDs and the downstream ECal. Image taken from [35].

2.3.2. Fine Grained Detectors

The fine grained detectors (FGD) act as the active target volumes for neutrino interactions in ND280 [47]. There are two FGD modules inside ND280, one constructed from scintillator bars and the second from scintillator bars and passive water targets. This configuration allows measurements on water as well as carbon. The FGDs are designed so that most of the particles produced from a neutrino interaction inside the module will reach the adjacent time projection chamber where their charges and momenta can be

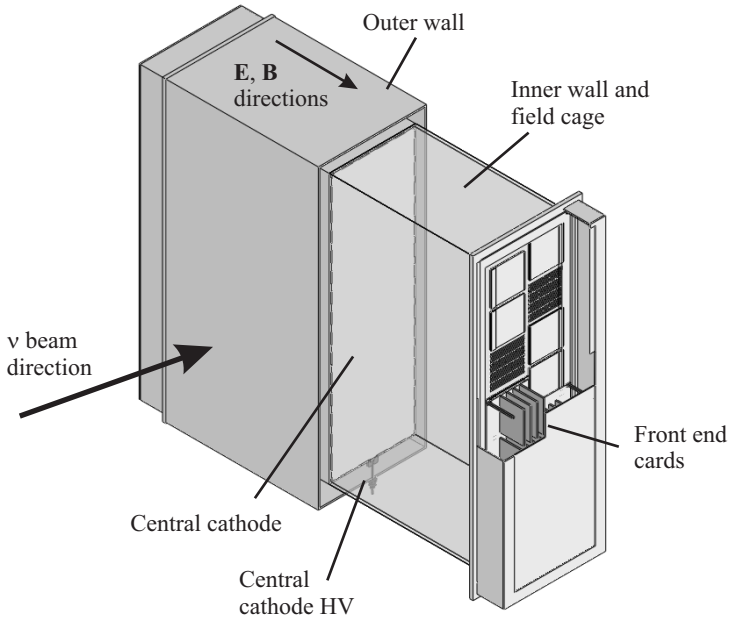


Figure 2.11.: Time projection chamber breakdown. Image taken from [48].

measured. However, shorter-ranged particles such as recoil protons are fully contained by the FGD and so the FGD has a fine grained resolution so that these particles may be tracked.

2.3.3. Time Projection Chambers

The time projection chambers (TPC), together with the FGDs, form the tracker component of ND280 [49]. The TPCs compliment the FGD by providing charged particle tracking and particle identification (PID) capabilities. Due to their high granularity and good momentum resolution, they can determine the track multiplicity and momentum of charged particles that traverse the detector and in addition to this can also be used to discriminate between charged particle types by measuring the ionization per unit length as shown in figure 2.12.

There are three TPC modules inside ND280 which contain 3000 l of gaseous argon in their volumes. A charged particle that passes through the argon gas will cause ionization, which is drifted with an electric field E towards bulk micromegas detectors which both amplify the charge and track the particle. A diagram of the TPC components can be

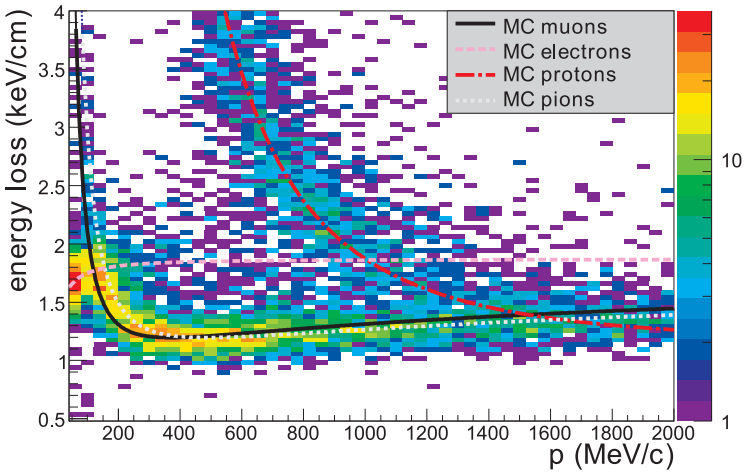


Figure 2.12.: MC and data comparison of TPC dE/dX energy deposition for positively charged particles produced in neutrino interactions. Figure taken from [48].

seen in Figure 2.11. The direction of the E field is aligned with the B field of the magnet, so that the drifted charge is not deflected by the magnet.

An example of an event inside the tracking detectors can seen in Figure 2.10.

2.3.4. Electromagnetic Calorimeter

The electromagnetic calorimeters (ECal) line the magnet and enclose the inner detectors [50]. There are 12 barrel modules which form a clam shell structure around the tracker and π^0 detector, and a single downstream (Ds) ECal module that sits at the end of the tracker detector section. Each module of the ECal is composed of scintillator bar layers stacked perpendicularly and inter-spaced with lead sheeting. Each bar is instrumented with MPPCs, either single or double ended depending on the module orientation. Each bar has a wavelength shifting fibre down the centre, which couples to the MPPC. If the bar has only one MPPC, the un-instrumented end has a mirror coating to reduce light loss. The ECal compliments the tracker detectors by providing energy measurements of $e^{\pm}$ and γ, additional particle PID, reconstruction of neutral particles and high-angle tracks with respect to the beam direction.

Further discussion on the ECal can be found in Chapter 3.

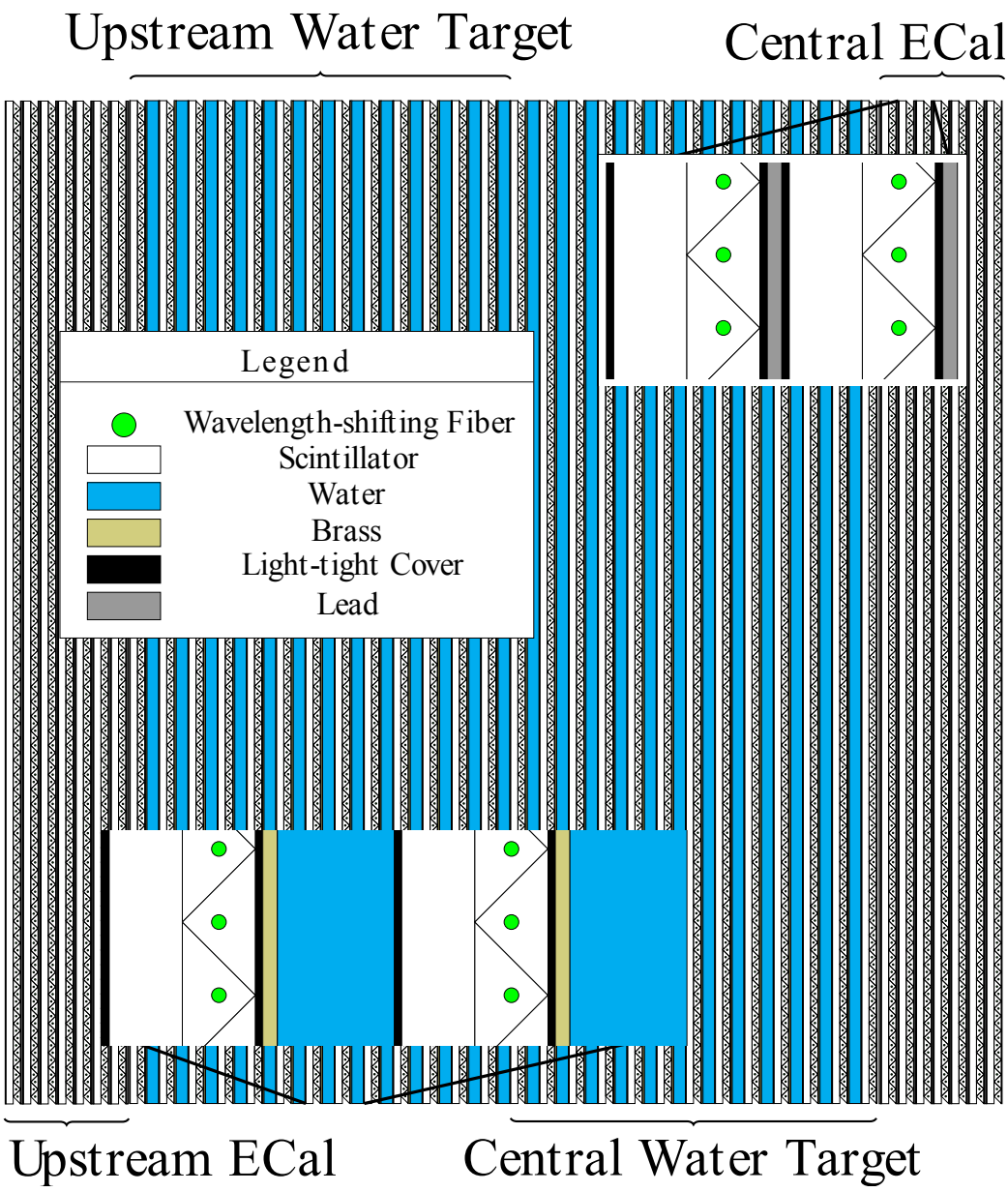


Figure 2.13.: Schematic of the P0D. The lead layers are placed upstream and downstream to act as a calorimeter, whereas the less massive brass and water constitute most of the layers, and act as targets for neutrinos. Image taken from [51].

2.3.5. π^0 Detector

The π^0 Detector (P0D) sits upstream of the tracking detectors. The P0D was designed to measure cross-sections for neutrino interactions involving π^0 particles [51]. As the dominant background for the ν_e appearance signal at SK was π^0 particles at the time of construction, the P0D was intended to be used to reduce the background by measuring

the properties of π^0s produced by the neutrino beam before it oscillates. However, recent developments in SK reconstruction, which are described in Section 6.3.2, have reduced the π^0 background. The P0D is built from triangular extruded plastic scintillator layers, inter-spaced with water bags, brass and lead as shown in Figure 2.13. The scintillator bars are read out via MPPCs.

2.3.6. Side Range Muon Detector

The Side Range Muon Detector (SMRD) is the outer most subdetector in ND280 and serves several purposes [52]. It is composed of extruded scintillator bars with WLS fibres laid into an S-shaped groove down the centre. The 440 scintillator bars that make up the SMRD are inserted into the air gaps in the UA1 magnet flux return yokes.

The SMRD provides ND280 with a cosmic trigger, as well as triggering on interactions that come from outside of ND280 and inside the magnet that occur within the beam spill trigger. Finally, the SMRD can assist in measurements of high angle muons coming from within the tracker.

2.4. On-Axis Near Detector (INGRID)

The Interactive Neutrino Grid (INGRID) detector sits on the neutrino beam axis and was designed to measure the beam direction and intensity [53]. They are of similar construction to the ECals (section 2.3.4) in that they are built from extruded scintillator bars interspaced with iron sheeting. Each bar is instrumented with a WLS fibre and an MPPC. As seen in Figure 2.14, INGRID is made up of 16 identical modules, 14 of which are arranged in a cross configuration, with the centre module sitting in the neutrino beam centre. The total mass of iron in each module is 7.1 t and at full neutrino beam intensity provides enough statistics to allow for daily updated measurements. The event rates in each module can determine the beam centre position to a precision better than 10 cm.

Additionally, two extra modules sit in between the two axes of the cross which monitor the axial symmetry of the beam.

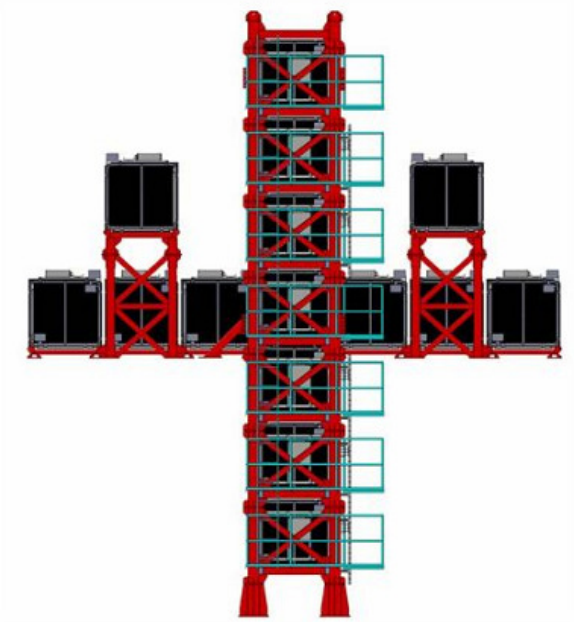


Figure 2.14.: Diagram of INGRID, showing the xy plane configuration of the modules. Image taken from [35].

2.5. Far Detector (Super-Kamiokande)

Super-Kamiokande (SK) is a 50 kt water Čerenkov detector situated 1000 m underneath Mt Kamioka, located in Gifu prefecture [54]. SK has optically isolated inner and outer detector volumes which share the same 50 kt volume of pure water (22.5 kt fiducial) and are lined with 11146 and 1885 photomultiplier tubes (PMT) respectively. An illustration of the SK detector can be seen in Figure 2.15. The outer detector (OD) encloses the inner detector (ID) and serves as a cosmic ray veto, and also to shield interactions from the neighbouring rock. The ID is used to detect neutrino interactions by measuring Čerenkov radiation from charged muons and electrons traversing the volume. The ID is a cylindrical volume of radius 33.8 m and height 36.2 m. The OD is larger than the ID in radius, top and bottom by an extra 2 m. An example of muon- and electron-like signals can be seen in Figures 2.16 and 2.17.

2.5.1. Čerenkov Radiation

SK's principle detection mechanism is via the imaging of Čerenkov radiation produced by ultrarelativistic charged particles traversing the water. When a charged particle travels

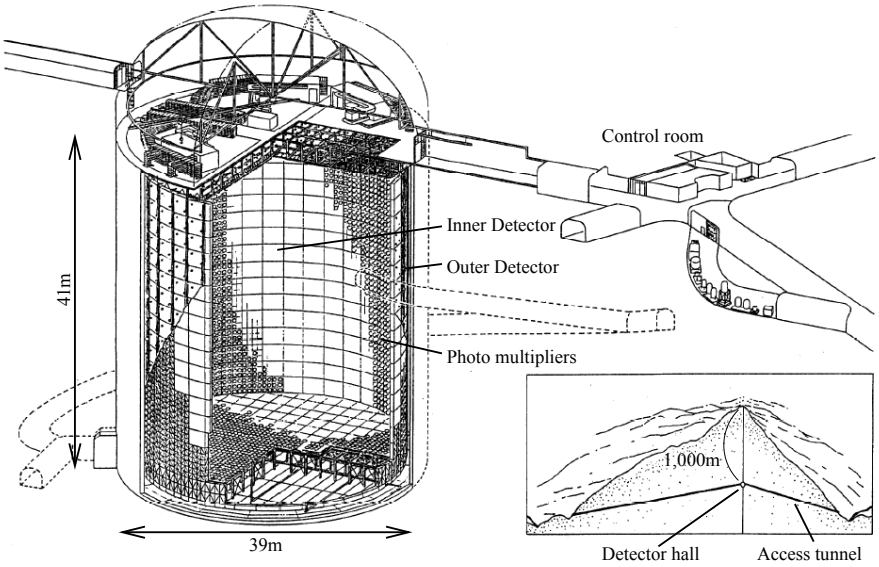


Figure 2.15.: Schematic drawing of the SK detector. The diagram is cut away to reveal the inside of the half-filled detector volume, which is filled with ultra pure water. The ID and OD share the same water but are optically separated. Diagram taken from [35].

through a medium with a refractive index n at a velocity v that is faster than the relative speed of light within the medium, the particle will create a cone of light emitted around the particle trajectory. The half-opening angle of the Čerenkov cone θ is related to the refractive index by

$$\cos\theta = \frac{1}{\beta n} \tag{2.9}$$

where $\beta = v/c$. These rings are imaged on the PMTs lining the detector walls. This light combined with timing information allows SK to reconstruct the neutrino vertex, neutrino direction and energy of the charged lepton.

Muons produce sharp rings of Čerenkov light, whereas electrons and photons shower, producing many overlapping rings which appear as a diffuse ring pattern (see Figures 2.16 and 2.17). Discrimination between these two types of rings is the basis of producing the muon and electron samples for the oscillation measurements. Each reconstructed ring undergoes particle identification (PID) by comparing the distribution of deposited charge to both analytically and Monte Carlo derived expectations for muon- and electron-like patterns respectively.

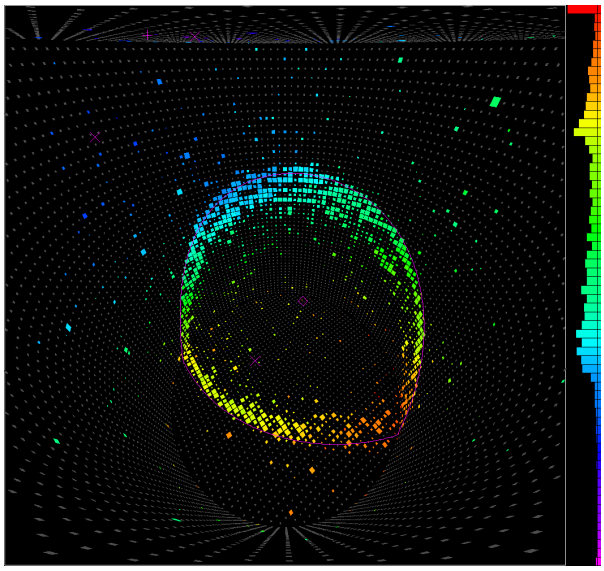


Figure 2.16.: Čerenkov light produced by a muon. The muon emits sharp rings of light as it travels through the ultra pure water. Image taken from [55].

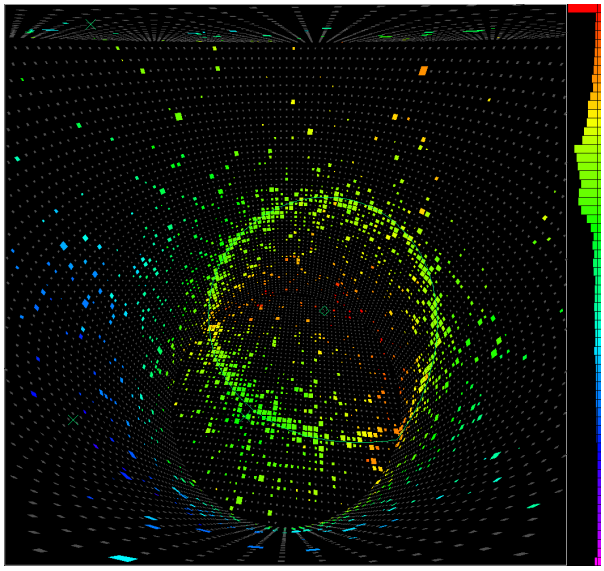


Figure 2.17.: Čerenkov light produced by an electron. The electron creates diffuse rings due to scattering of the electron on the medium. Image taken from [55].

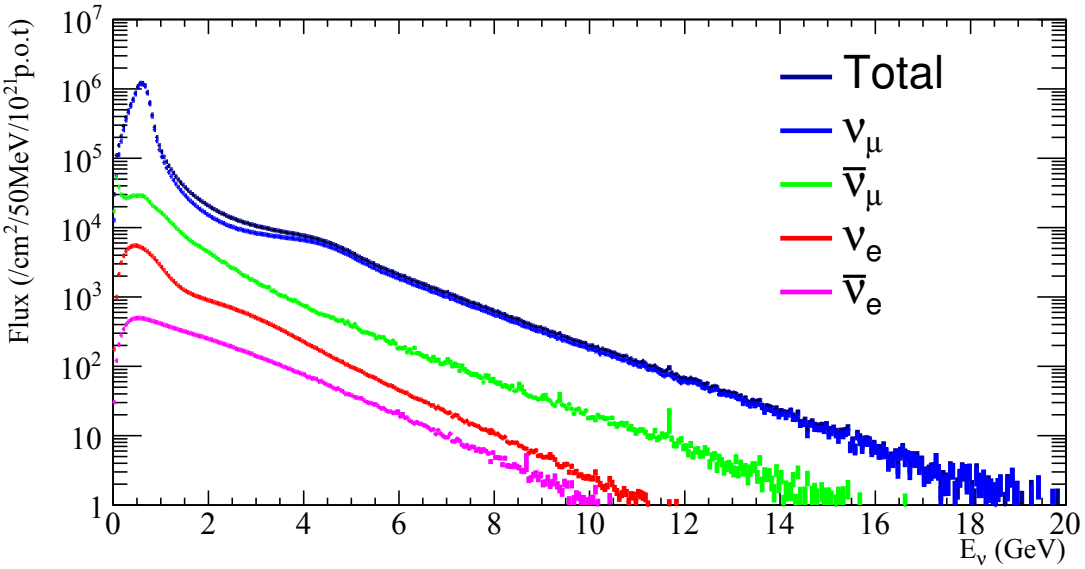


Figure 2.18.: Flux prediction at SK.

2.6. Simulation of T2K

The T2K experiment simulates the neutrino beam from the point where the proton beam interacts with the target, all the way to the primary and secondary interactions inside the detectors. This is necessary to have a complete understanding of the experiment and to produce Monte Carlo (MC) for the oscillation and cross-section measurements. The neutrino generator tools NEUT [56] and GENIE [57] are used.

2.6.1. Flux Simulation

The simulation of the neutrino flux starts with the proton beam interacting with the target [59]. External data from the NA61/SHINE experiment [58] is used to reweight the hadronic interactions of beam with target and baffle produced with FLUKA [60]. An example of this reweighting can be seen in Figure 2.19. Primary particles produced in this process are propagated through the secondary beamline (see Figure 2.7b) with GEANT3 [61] and secondary interactions are simulated with GCALOR [62]. The neutrino flux is estimated from the simulated kinematics of the particles that decay in the decay volume. The flux prediction at SK can be seen in Figure 2.18.

Figure 2.19.: Flux prediction at ND280, as predicted by FLUKA (left) and reweighted by NA61 data (right). Flux is broken down by the parent particle type. Figures taken from [58].

2.6.2. Detector Simulation

The ND280 geometry is simulated with GEANT4 [63], and interactions are handled by GENIE and NEUT. The simulation of final state particles in ND280 are performed with GEANT4. The scintillator, MPPC and electronics response are simulated with custom written software elecSim. Interactions in SK are simulated with event generators GENIE and NEUT. Detector response simulated by SKDETSIM, a package that interfaces with GEANT3.

2.6.3. Neutrino Interaction Simulation

The interactions of neutrinos inside the detectors in T2K are simulated by default with NEUT; GENIE serves as an alternative generator.

Several models are implemented in NEUT to simulate different ways the neutrino can interact. For CCQE interactions, the Llewellyn-Smith model [64] is used to model the neutrino quasi-elastic scattering with a nucleon. Resonant pion production is modelled using the Rein and Sehgal model [65] which considers decays of resonant 3 quark systems. Coherent pion production is modelled by simulating the neutrino coherently scattering off the whole nucleus [66]. Deep inelastic scattering modelling uses nucleon structure

functions [67] and the Woods-Saxon model to describe the nucleon density distribution for particles travelling through the nucleus [68].

Chapter 3.

The Barrel Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECal) modules of the ND280 detector line the magnet and surround the basket, to contain most particles that are produced within the tracker and P0D detectors. The barrel Electromagnetic Calorimeter (BrECal) modules surround the tracker region (as shown in Figure 2.8) and capture high angle leptons, measure electron and positron showers to improve energy measurements, detect neutral particles and can veto cosmic muons that traverse the detector [69]. The downstream ECal (DsECal) sits inside the basket downstream of the tracking detectors.

This chapter discusses the BrECal composition and its construction, and a description of the online monitoring and commissioning of all ECal modules.

The ECals main detection mechanism is by inducing electromagnetic showers from $e^{\pm}$ and photons. The ECals are sampling calorimeters, meaning they are segmented; the active components are plastic extruded scintillator bars layered between lead sheets. When an electron or gamma reaches the lead layer, showering can occur which causes a cascade as shown in Figure 3.1. Showering occurs when an $e^{\pm}$ or photon enters the lead and will emit bremsstrahlung, or pair produce respectively. The $e^{+}e^{-}$ pair will scintillate inside the scintillation bar, and go on to produce more showering. The photon emitted by bremsstrahlung will itself undergo pair production, thus the main showering mechanism for both $e^{\pm}$ and photons is essentially the same; the main difference is the initial interaction, and the fact that the photon is undetected until it showers.

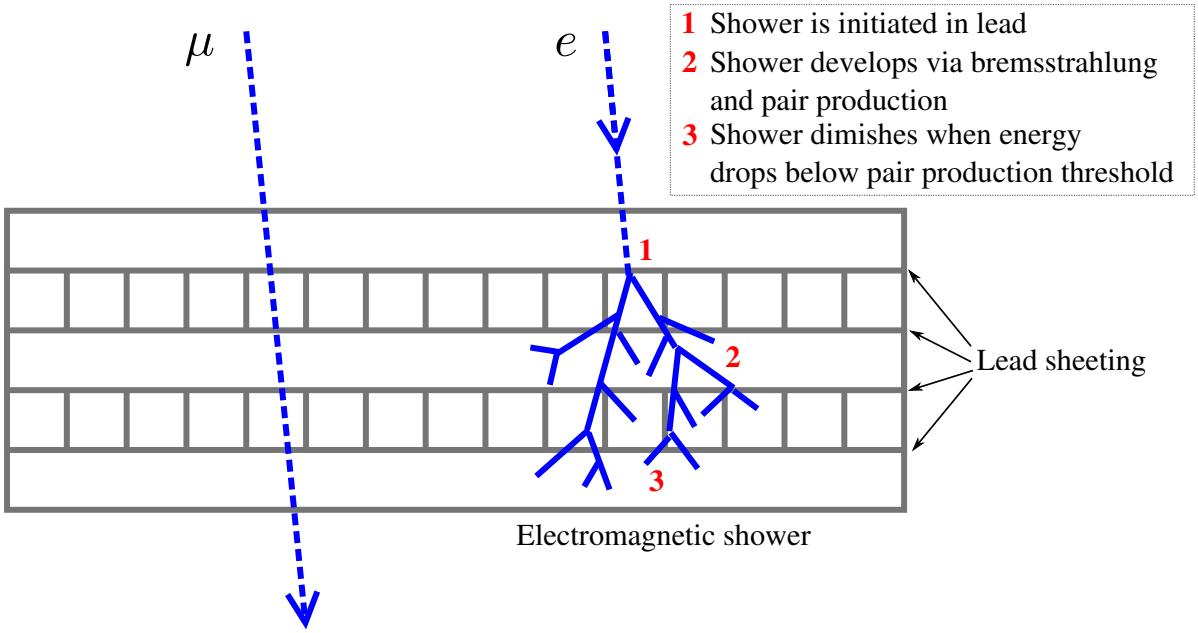


Figure 3.1.: Diagram showing the different mechanisms for energy to be deposited in the ECal. $e^{\pm}$ and photons will start to shower in the lead layer, which produces scintillation light. A minimum ionizing particle such as a muon will not shower, but will cause fluorescence in the scintillation medium.

3.1. BrECal Components

The main purpose of the BrECal is to improve the measurements of CCQE events by measuring the high-angle leptons leaving the tracker [70]. This involves discriminating between muon-like tracks and electron-like showers. The BrECal is composed of 6 separate modules, which are shown in Figure 3.2.

The two side modules are the largest as they span the full height of the tracker. The top and bottom faces of the tracker have two ECal modules each which allow the magnet to open and close. The inward facing side of the top and bottom modules are uninstrumented. Table 3.1 contains details of the modules.

3.1.1. Fibre Optics

The ECal modules have Kuraray Y11 wavelength-shifting (WLS) fibres [71] threaded through each scintillator bar. The WLS fibres changes the wavelength of the scintillation light from UV to the green region of the spectrum, which is better suited for detection

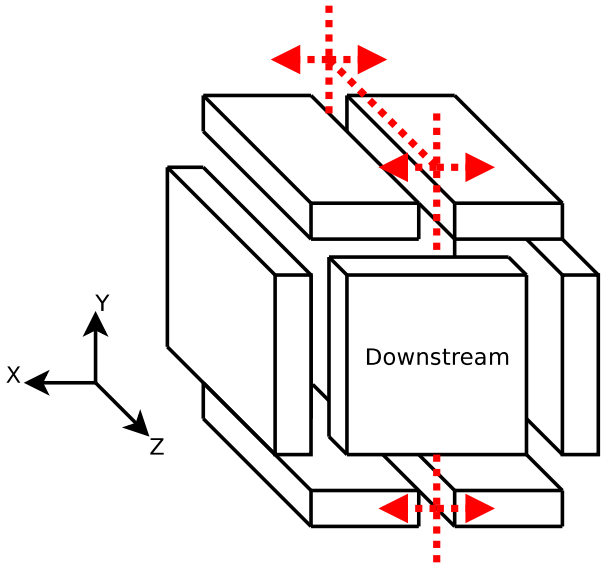


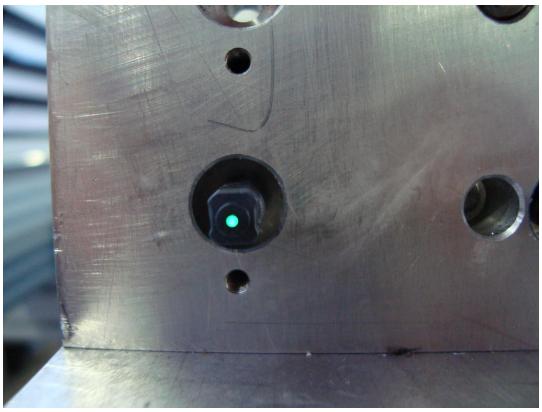
Figure 3.2.: Configuration of the Barrel and Downstream ECal modules. The red arrows show how the Barrel modules move when the magnet is opened. The Downstream module is held within the basket and doesn't move. Not shown are the P0D ECal modules, which sit upstream of the Barrel modules and surround the P0D in the same way that the Barrel surrounds the tracker.

	Barrel Side	Barrel Top / Bottom	Downstream
Layers	31	31	34
Dimensions (m)	4.14×2.503	4.14×1.676	2.3×2.3
Bars per long	57	38	50
Bars per short	96	96	50
Instrumentation	double-ended	short bars single-ended	double-ended

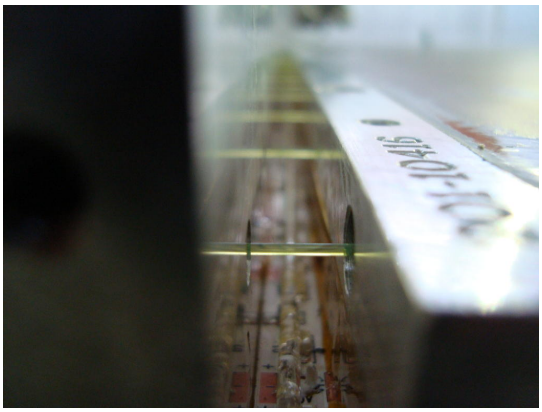
Table 3.1.: Details of the barrel and downstream ECal modules. The dimensions quoted are the relative xy plane of each module; the height is neglected as it can be inferred from the number of layers. The number of bars are quoted for long and short layers, referring to the length of the bars used in each layer, which are stacked perpendicularly. Instrumentation refers to whether the bars have an MPPC coupled to both ends or not.

by the MPPCs. The WLS fibres and plastic ferrules that couple the fibre to the MPPC can be seen in Figure 3.3.

The WLS fibres were manufactured in Japan, and were then shipped to Fermilab to be cut into correct lengths with a tolerance of ± 0.5 mm and have the fibre ends



(a) Ferrule and fibre face protruding from the bulkhead



(b) Fibre exposure between layer and bulkhead

Figure 3.3.: Images showing the fibre optic elements of the ECal.

ice-polished. The ice-polishing consists of supporting batches of fibres using ice, and then polishing the ends using diamond. Quality assessment was performed at the University of Warwick where the plastic ferrules were also glued. Lastly, they were sent to Liverpool and Daresbury for installation.

The top and bottom modules that have an uninstrumented side have only single ended WLS fibres; a mirror coating is applied to one end. All other bars have an MPPC attached to both ends. The insertion of the fibres into the double-ended bars is more complicated, as the fibre must have a ferrule glued to one end after installation using BC-600 Optical Cement. BC-600 was chosen to not alter the optical properties of the fibre, and consists of a resin mixed with a hardener at a ratio of 2:0.56 g which is then left for $2\frac{1}{2}$ hours before it can be used.

A compressed air glue gun is used to apply short bursts of glue to the fibres. The ferrules are then placed onto the ends of the fibres, and are held in place with a spring mounted cap that is temporarily screwed to the bulkhead. The fibre should protrude approximately 0.2 mm from the ferrule in order to minimise light leakage between the ferrule and MPPC.

3.1.2. Scintillation Layers

Each layer must be fully constructed before it is inserted into the module. A single layer consists of a metal frame that has holes milled for fibres to be inserted, which surrounds

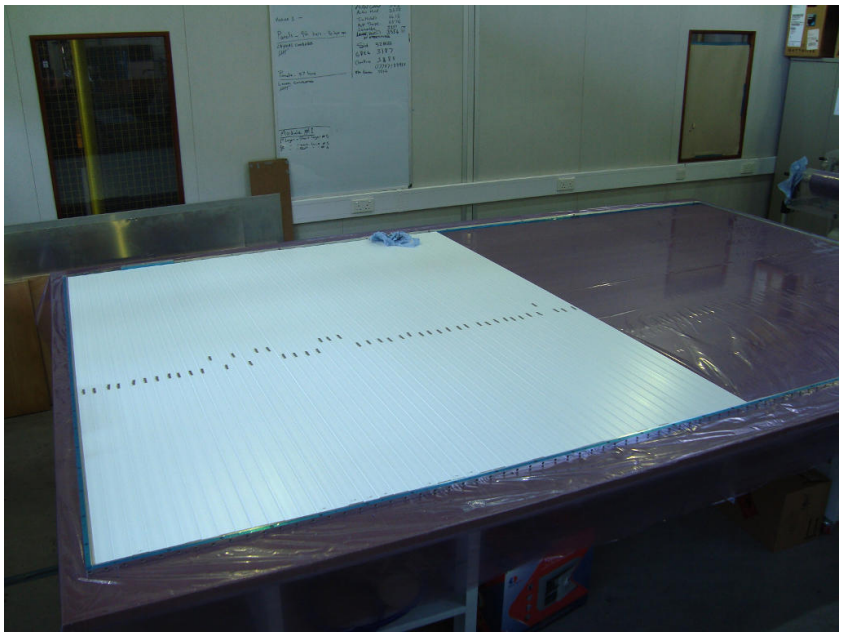


Figure 3.4.: A layer being assembled at Daresbury Laboratory. The scintillator bars are in the process of being inserted into the metal frame.

an array of scintillator bars (seen in Figure 3.4), which are glued into place and topped with a layer of lead sheeting 1.7 mm thick.

The scintillator bars are made from extruded polystyrene doped 1% POPOP ($C_{24}H_{16}N_{2O}$) and 0.03% PPO ($C_{15}H_{11}NO$). A photon causes an excitation in the polystyrene, which is transferred into the PPO molecule which emits UV light. This UV light is wavelength shifted by the POPOP to improve the transmittance. The extrusion method used to construct the bars allows for a precision mixture of the three materials, as well as the creation of a central hole for the fibre to sit in, which would be impractical to drill due to the bar length. The bar dimensions are 1×4 cm with variable length depending on which module they belong to. The scintillator bars are coated with titanium oxide to promote internal reflection of light.

3.1.3. Light Injection System

The light injection (LI) system is a pulsed light source that is designed to be used for timing calibration of the MPPCs. LED strips are placed between the bulkhead and first layer, which illuminates the exposed WLS fibres as seen in Figure 3.3b.

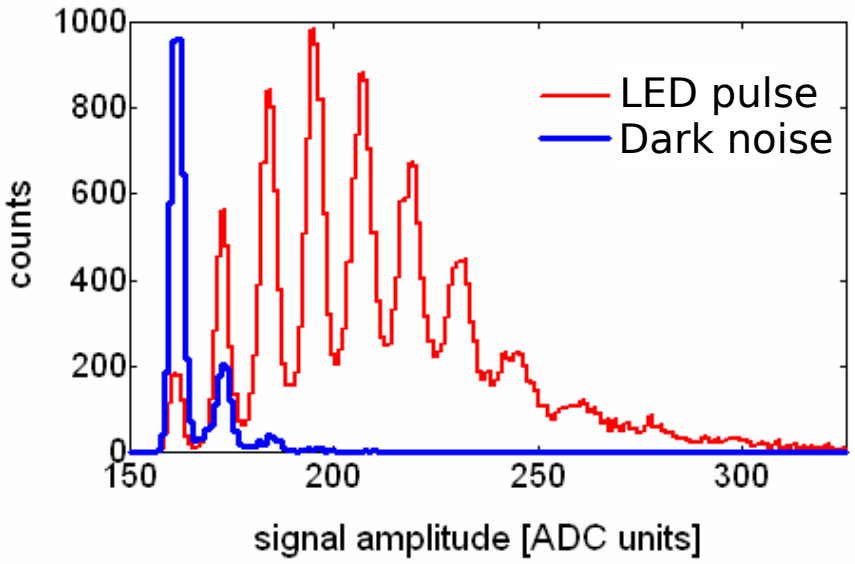


Figure 3.5.: A dark noise signal compared to a signal produced by an LED pulse. Figure taken from [73].

3.1.4. Electronics

The light from the WLS fibre is incident upon the face of the MPPC coupled to it. The MPPCs connect to Trip-t frontend boards [72] (TFBs) via a mini-coaxial cable. Each TFB has 4 Trip-t integrated circuits which have 16 channels each. In total each TFB can instrument up to 64 MPPCs. The Trip-t chip contains a capacitor bank of 23 capacitors, which stores charge from the MPPCs. These capacitors are integrated in readout cycles that synchronise with the timing signals from the neutrino beam. An example of MPPC photoelectron spectra can be seen in Figure 3.5.

Each TFB connects to a readout merger module [73] (RMM) via a CAT-5e twisted pair cable. The ECal has 12 RMMs which handle data readout and also issue clock and trigger signals to up to 48 TFBs each. RMMs are controlled by the master clock module (MCM) and also by the cosmic trigger module (CTM) which provides triggering information for cosmic rays in between neutrino beam spills. The RMMs send data to front end node machines which perform fast data processing tasks for online monitoring.

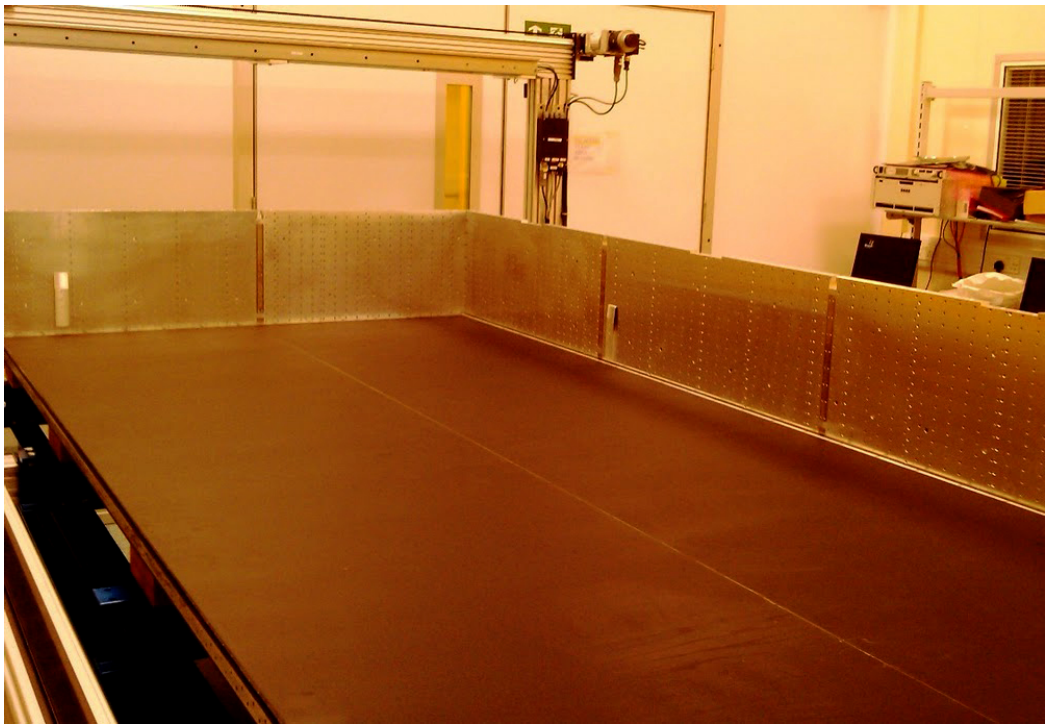


Figure 3.6.: Bulkheads attached to the module base, with the first layer inserted.

3.2. Construction of the Barrel Electromagnetic Calorimeter

Construction of the BrECal began in the Oliver Lodge Laboratory at the University of Liverpool and Daresbury Laboratory in July 2009, with each lab building a separate module of the detector concurrently. Construction was completed in late 2010.

Although the dimensions are different, the same process of construction is used to build each BrECal module. A carbon fibre base is built with an aluminium supporting frame, which is laid down onto a supporting structure. The carbon fibre provides strong support for the weight of the module, whilst having a low enough density to reduce interactions within the material by charged leptons.

Stainless steel bulkheads are screwed in place around the edges of the base. In the case of the top and bottom modules, one side is uninstrumented and has no bulkheads. The bulkheads consist of 1.27 cm thick aluminium with an array of holes drilled into the sides to which the MPPCs are attached. The bulkheads are the main structural support

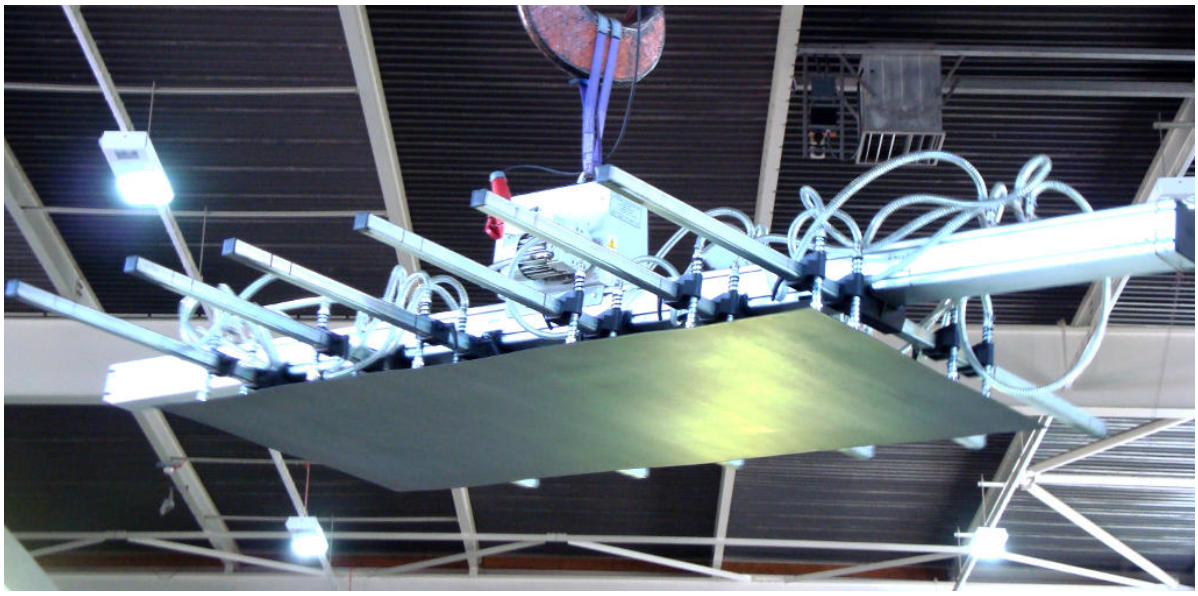


Figure 3.7.: Lead sheet being lowered into position using a vacuum lift at Daresbury. The vacuum lift is also used to drop layers into the module.

for the module and can be seen in Figure 3.6. The first layer of scintillator bars is then inserted, after which the light injection system can be installed.

The layers are built individually by ordering plastic scintillator bars within a metal frame and fixing them in place with Araldite epoxy adhesive. Depending on the module, the layer will have differing numbers of scintillator bars and orientations. Lead sheeting is then placed on top of the layer using a vacuum lift as shown in Figure 3.7, and affixed by Araldite epoxy. The layer is covered in plastic sheeting which is evacuated in order to let the adhesive dry with no air bubbles.

Once dry, the layers are lifted into the module framework and screwed down. Then wavelength shifting fibres are threaded through the bulkhead holes and into the layers. MPPCs are then coupled to the fibre via a plastic ferrule and attached to the bulkhead.

Upon completion of a layer, the bars are validated via scanning the bars with a ^{137}Cs radioactive source and measuring the signal in each bar. This procedure is detailed in Section 3.3. When the layer is validated, the next layer is installed and the process repeated until all layers are inserted. A module with all layers installed can be seen in Figure 3.8.

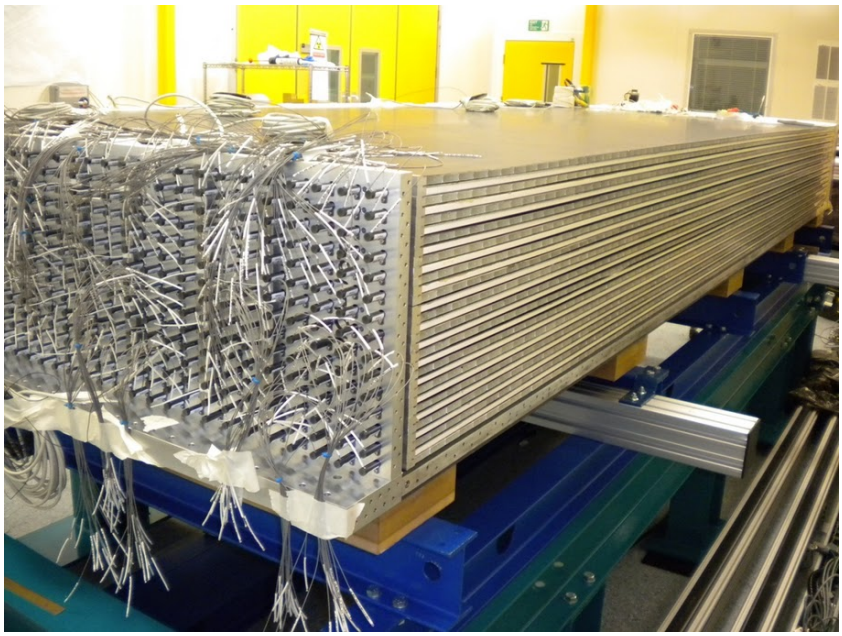


Figure 3.8.: A module with all layers installed, and with all MPPCs attached to the bulkhead.

After the MPPCs are attached, the TFB electronics are attached to cooling plates, which are mounted in front of the bulkheads as seen in Figure 3.9. The coaxial cable which connects the MPPC to the TFB is threaded through holes in the cooling plates.

Finally, power supply lines and CAT-5E cables for data transfer are installed as shown in Figure 3.10. The module then has an outer casing and metal lid attached, and then is ready to be shipped to the ND280 complex in Japan.

3.3. Validation of Scintillator Bars

The fibre integrity of the bars are validated at the time of installation by scanning the newly installed layer with a ^{137}Cs gamma source to induce a signal inside the bar. A set of well understood MPPCs were used to check every bar, connected to two TFBs and one RMM as shown in Figure 3.11. The gamma source is positioned at many points along each bar, and the data is recorded. The attenuation profile follows a double exponential function

$$A_1 e^{x/\lambda_1} + A_2 e^{x/\lambda_2} \tag{3.1}$$

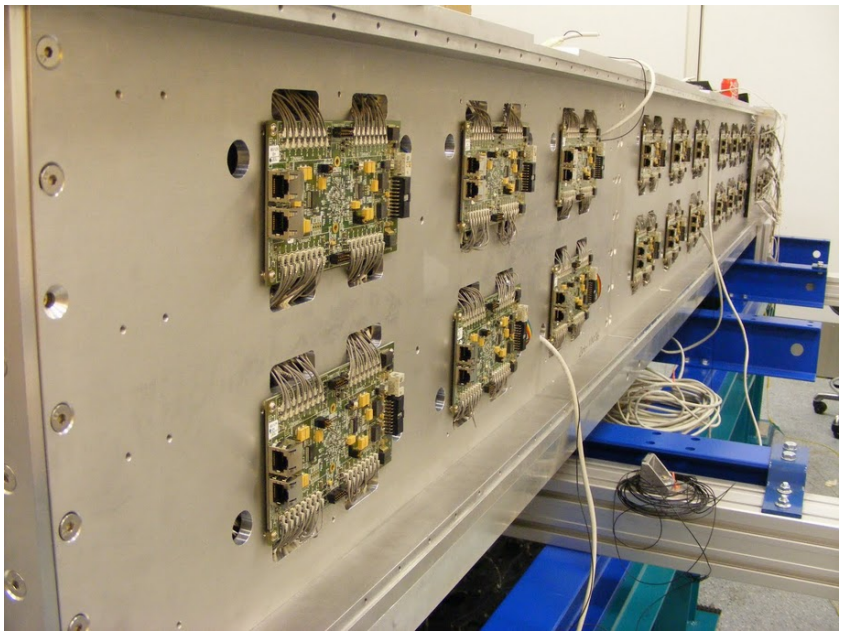


Figure 3.9.: Cooling plates with TFB electronics attached. The coaxial cable from the MPPCs are threaded through gaps in the cooling plates, and connected to the corresponding channel on the TFB.

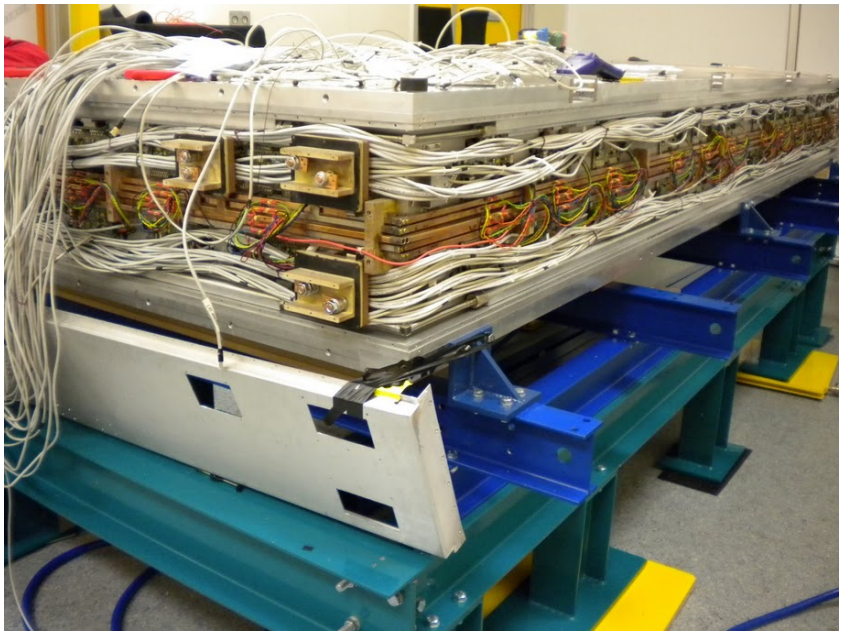


Figure 3.10.: All components of the module assembled.

as shown in Figure 3.13. Further attenuation of the signal at the bar ends is caused by edge effects. Typically, an unhealthy fibre has a sudden drop in attenuation, signalling a break in the fibre. An example of this can be seen in Figure 3.14. A break in the fibre could occur during transportation, or when being handled during construction.

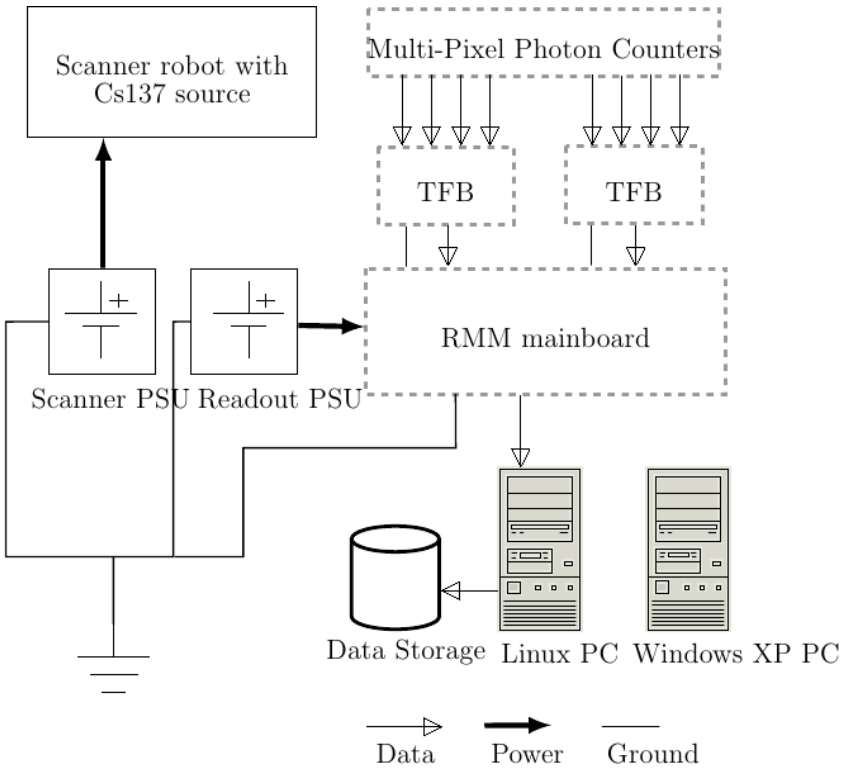


Figure 3.11.: Diagram showing the set-up of electronics for the bar validation scanning procedure.

Scanning Process

To produce gamma rays, a ^{137}Cs source is used, which decays to ^{137}Ba, that finally decays emitting a 0.6617 MeV photon. This gamma ray undergoes Compton scattering off of the electrons in the lead atom, which induces a signal inside the scintillator bar.

Upon starting the scanning process, the scanner arm will collect the source from a secure docking station, and proceed to the first bar. An image of the source being scanned across a layer can be seen in Figure 3.12. The scanner will expose the bar to gamma radiation at various positions along the bar (see Figure 3.15), building the attenuation profile. It typically takes 5 hours to complete the scan of a whole layer, which produces ADC spectra for each MPPC channel.

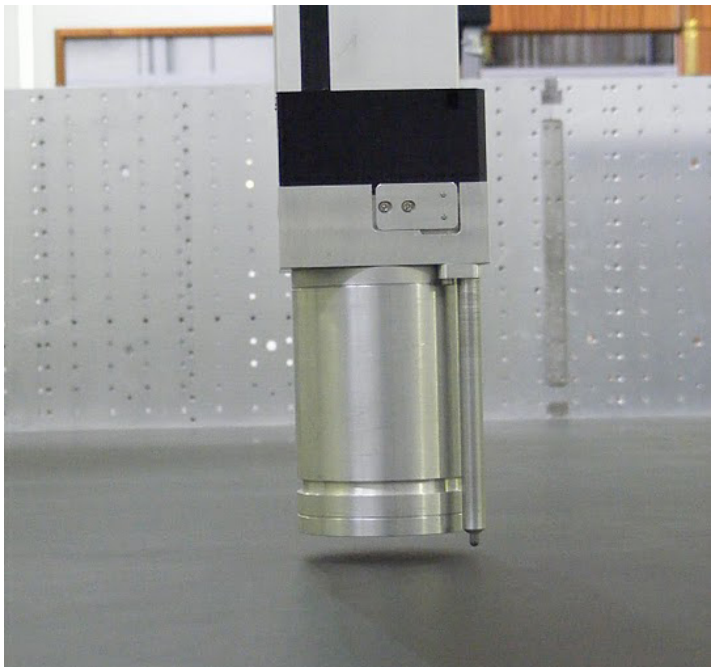


Figure 3.12.: Image of the radioactive source attached to the scanner at Daresbury. The scanner has a probe that senses and prevents the scanner arm from damaging the layer.

To reduce noise in the photosensors, the sides of the module are covered with a light tight material. Any remaining light leaks are found by recording dark noise data, and visually inspecting the MPPC photoelectron spectra.

3.4. Calibrating MPPC Gain

There are on the order of 23,000 channels instrumented with MPPCs in all the ECal modules. The variation of temperature (both seasonal and daily) will affect the response of each MPPC, as the breakdown voltage is temperature dependent as seen in Equation 2.8. This means that the performance of the ECal must be monitored daily, and frequently recalibrated to ensure optimum signal to noise ratio of each channel.

3.4.1. Calibration Procedure

Each MPPC produces a photoelectron spectrum, which consists of a series of peaks, as shown in Figure 3.16. Characterisation of this spectra is used to set the optimum gain

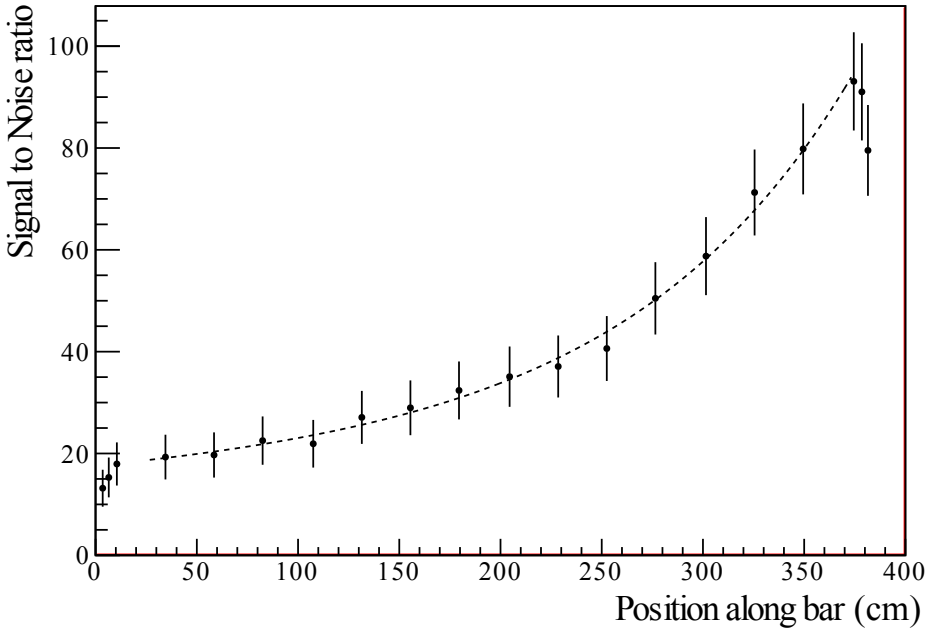


Figure 3.13.: Attenuation profile of a bar constructed from the scanning procedure.

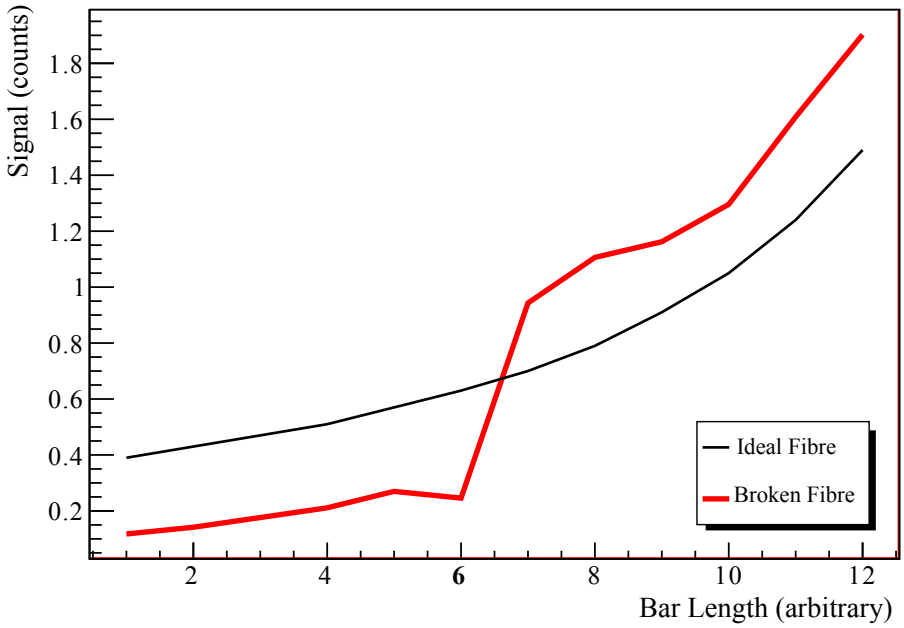


Figure 3.14.: Broken fibre compared to an ideal healthy fibre, constructed by averaging over many healthy fibres.

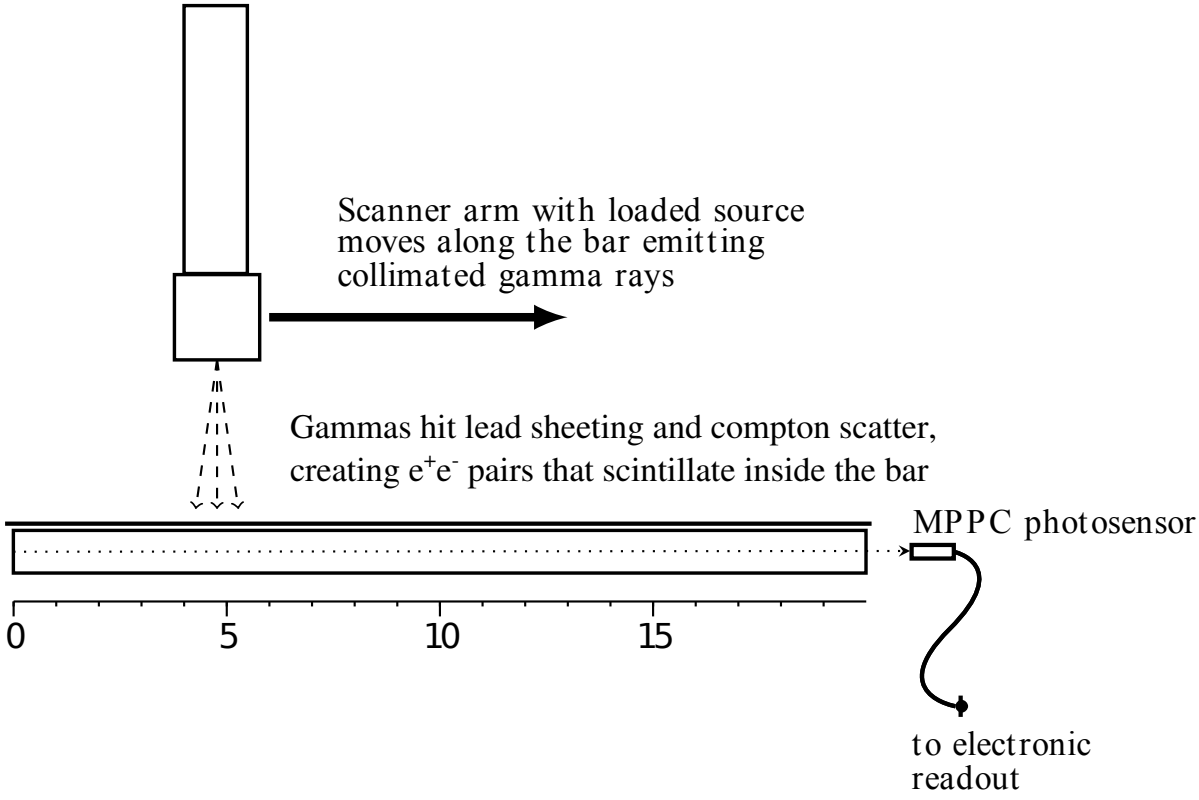


Figure 3.15.: Diagram showing the process of scanning a scintillator bar with the ^{137}Cs source.

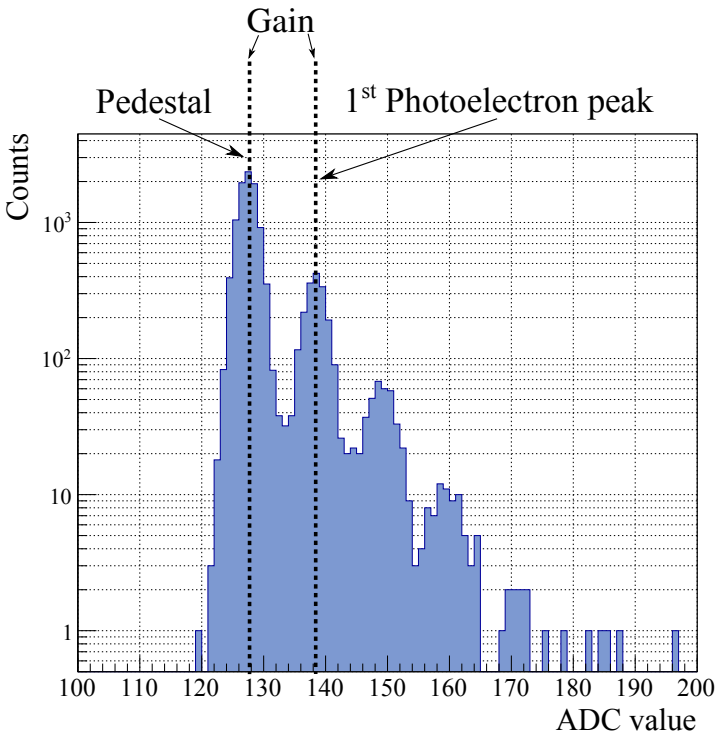
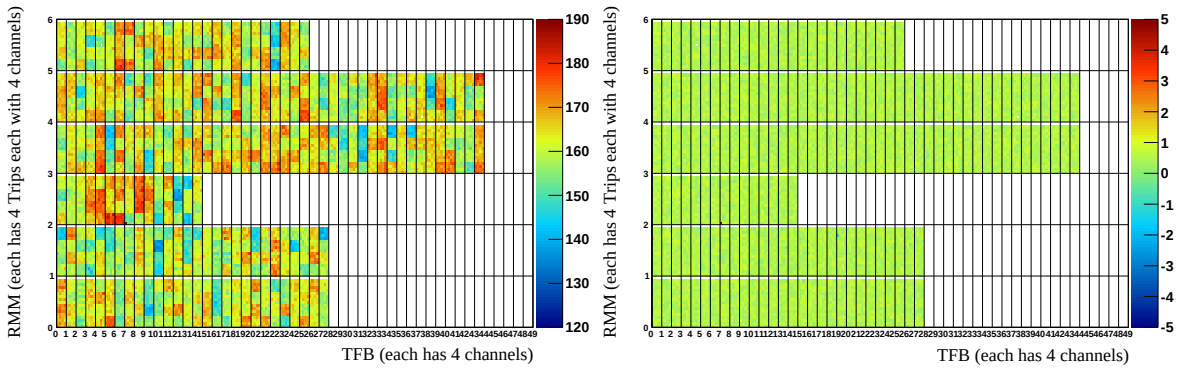


Figure 3.16.: A typical MPPC photoelectron spectra for a pedestal run.

of the device, which is found to be an overvoltage $V_{op} - V_{bd}$ =1.3 V, corresponding to approximately 14 ADC counts between the pedestal and first photoelectron peak.

The first peak in the MPPC photoelectron spectrum is referred to as the pedestal which is visible when there is no incident light. Its position is dependent on temperature and also the intrinsic properties of the device during the manufacturing process. The ECal is triggered at set intervals with no physics trigger so that dark noise is collected. The resulting pedestal data is used to monitor the state of the detector. A Gaussian fit to the pedestal for each MPPC is performed, and the pedestal position stored (see Figure 3.17a). This pedestal position is used to center the pedestal peak of each MPPC photoelectron spectra at 0, as in Figure 3.17b. Temperature induced drift of the pedestal position can be monitored and used to determine whether the detector requires recalibration.

To calibrate the gain, a series of pedestal runs for varying operational voltages V_{op} are taken. The gain of an MPPC is calculated by the difference in ADC counts between the pedestal and first photoelectron peak, which are obtained by a Gaussian fit to each peak. The gain is calculated for each channel, plotted as a function of V_{op}, and fit with a straight line. The optimum V_{op} is selected from this fit result for the value that corresponds to a gain of 14. Newly calibrated gains of each MPPC can be seen in Figure 3.18.



(a) Pedestal ADC positions. **(b)** Current pedestal position minus stored position.

Figure 3.17.: Per MPPC information, arranged by RMM and TFB groupings.

Each RMM instruments physically different regions of the ECal modules, and so observing the average gain of MPPCs per RMM, one can see how minor differences in environment can affect the MPPCs. Figure **??** shows the distribution of MPPC gain per RMM, for a set of V_{op} values that are no longer optimal due to temperature variation. Groupings of MPPCs appear to form distinct distributions, which reflect again the affect of temperature variation across the detector. Figure 3.19 show the gain distributions of two ECal modules before and after gain calibration. The left column show a set of V_{op} values that are no longer suitable, and the right column show newly calculated values. The top row corresponds to the DsECal, and the bottom row corresponds to a BrECal module. The DsECal is not mounted to the magnet, and so is more susceptible to variations in temperature. This is reflected in the fact that the variation in gain values is slightly wider than compared with other modules.

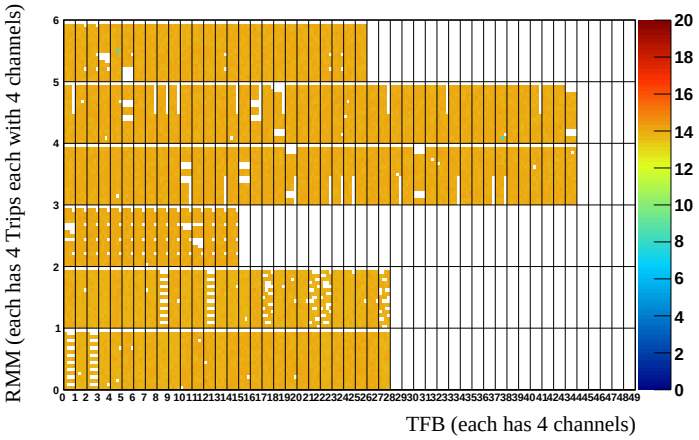


Figure 3.18.: Gain values for each MPPC, arranged by RMM and TFB groupings.

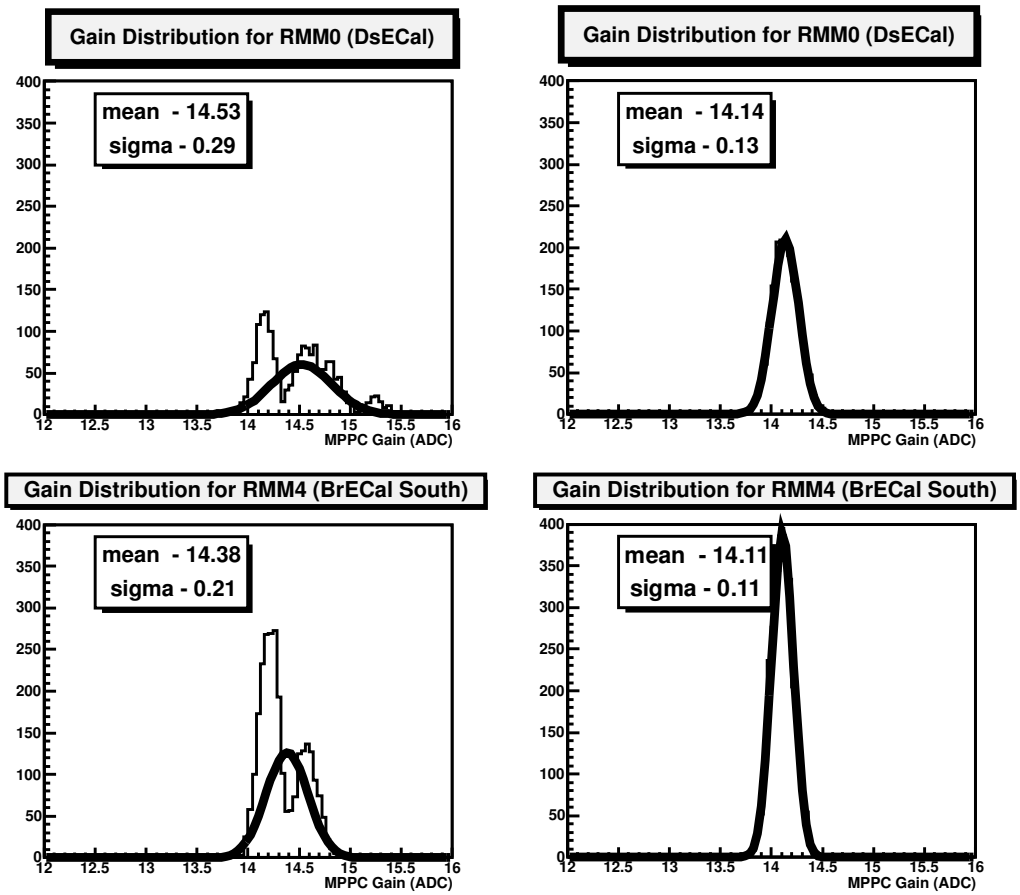


Figure 3.19.: Bad gain distributions per RMM. The distributions show large variations from the nominal gain value.

Chapter 4.

Development of a Bayesian Oscillation Analysis Framework

T2K oscillation analyses are composed of many self contained stages of analysis, which are due to the separate detector complexes. For example, a typical oscillation analysis will take the fit result of the near detector data, constrained by beam flux, detector and cross-section simulations, and extrapolate this result as an input to the far detector analysis [74].

The motivation to develop a modular framework whereby multiple samples from different detectors may be combined together in a common analysis is strong, in that there are strongly correlated systematic uncertainties between the near and far detector, along with complimentary ν_μ and ν_e samples at the far detector.

Such a framework has two main requirements:

- be modular enough to expand to any number of samples and high numbers of systematic parameters
- provide a consistent treatment for correlated systematic uncertainties between data samples.

This chapter details the efforts to develop an analysis framework that satisfies these requirements.

The issue of modularization has been addressed by writing base classes which systematic models and detector Monte Carlo samples can inherit from and interface to the main fitting mechanism. The main fitting mechanism utilized is a Markov chain Monte Carlo (MCMC); a description of MCMC can be found in Appendix C, and it is advised

to be familiar with the key concepts before continuing. A MCMC analysis scales well to high numbers of parameters, and allows full marginalization of all nuisance parameters. MCMC has the added advantage of being able to run multiple chains and then combine the finished output. This makes it easy to distribute the computional load across a cluster.

Regarding the treatment of systematic uncertainty, the framework employs an event-by-event reweighting method to construct the probability density functions required in the likelihood calculation steps. This allows one to treat migrations between bins in a binned likelihood calculation, and also event category migrations. This is possible because all original MC information is retained, instead of using histogram templates. However, due to the orders of magnitude more event weight calculations that must be performed, a huge hit to fitter efficiency is taken. This is addressed by off-loading some of the large bottlenecking calculations to a Graphics Processing Unit (GPU) which are present in many modern systems (see Appendix D, Chapter 5 and Section 4.1.4).

The resulting analysis framework has been dubbed MaCh3, reflecting its utilization of a **Ma**rkov **ch**ain Monte Carlo for a **3** flavour oscillation analysis. Section 4.1 discusses the technical aspects of implementing the analysis framework and obtaining the posterior distribution, and Section 4.2 discusses the statistical methods of extracting information from the posterior relevant for parameter inference.

4.1. MaCh3 Framework

MaCh3 uses an MCMC to construct the posterior probability distribution (to be discussed in Section 4.2) of the 3 flavour oscillation and systematic uncertainty models in parameter space. The MCMC algorithm is illustrated in Figure 4.1. It is written so that samples and parameters are distinct objects, each handling their own reweighting and likelihood operations which can be controlled by a fitter. As seen in Figure 4.2, the MCMC algorithm is separated from the parameter and sample classes, which allows the analyser to combine multiple combinations of samples and parameters to create an analysis.

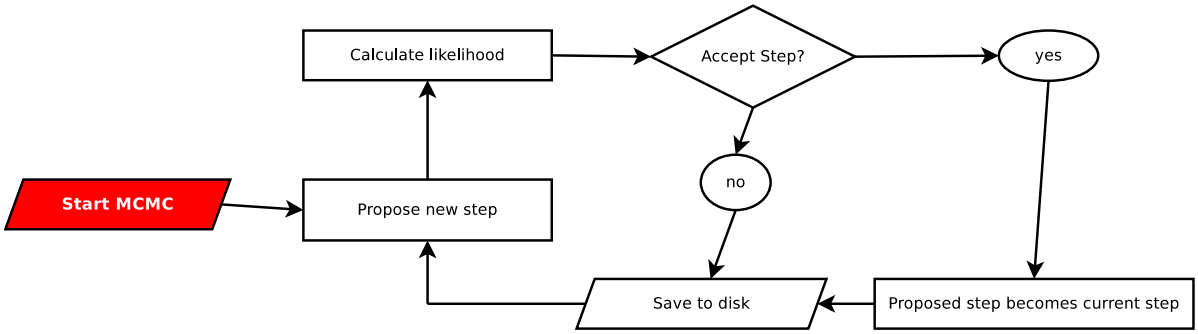


Figure 4.1.: MCMC algorithm flow chart.

4.1.1. Base Classes

MaCh3 is programmed in an object oriented style so as to promote modularity and flexibility of analyses in T2K. The main method of achieving this is through the use of base classes for all object types. The two main base classes are the sample base class and the systematic base class. These base classes contain the fundamental functions and data structures that all samples and systematic parameters will inherit from.

The sample base class includes data structures to store detector MC and methods for extracting, formatting and reweighting them. The systematics base class handles covariance matrices and the generation of correlated steps through Cholesky Decomposition (see Section 4.1.2).

Both the sample and systematics base class provide likelihood terms which are added piecewise by the MCMC class. Modularity is achieved through polymorphism; the MCMC class has a stack that holds both sample classes and systematics classes, and will iterate through them and request the likelihood terms from each. This enables samples and systematics to be combined in a "plug-and-play" manner.

4.1.2. Step Proposal

At each iteration of the MCMC a new state in parameter space must be proposed. This is done by sampling from a proposal function that meets the requirements set out in Appendix C. For larger, and especially highly correlated parameter spaces, the tuning of the proposal function is vital to achieve good chain mixing and thus decreasing the length of the chain required. A discussion of optimum proposal function choice can be

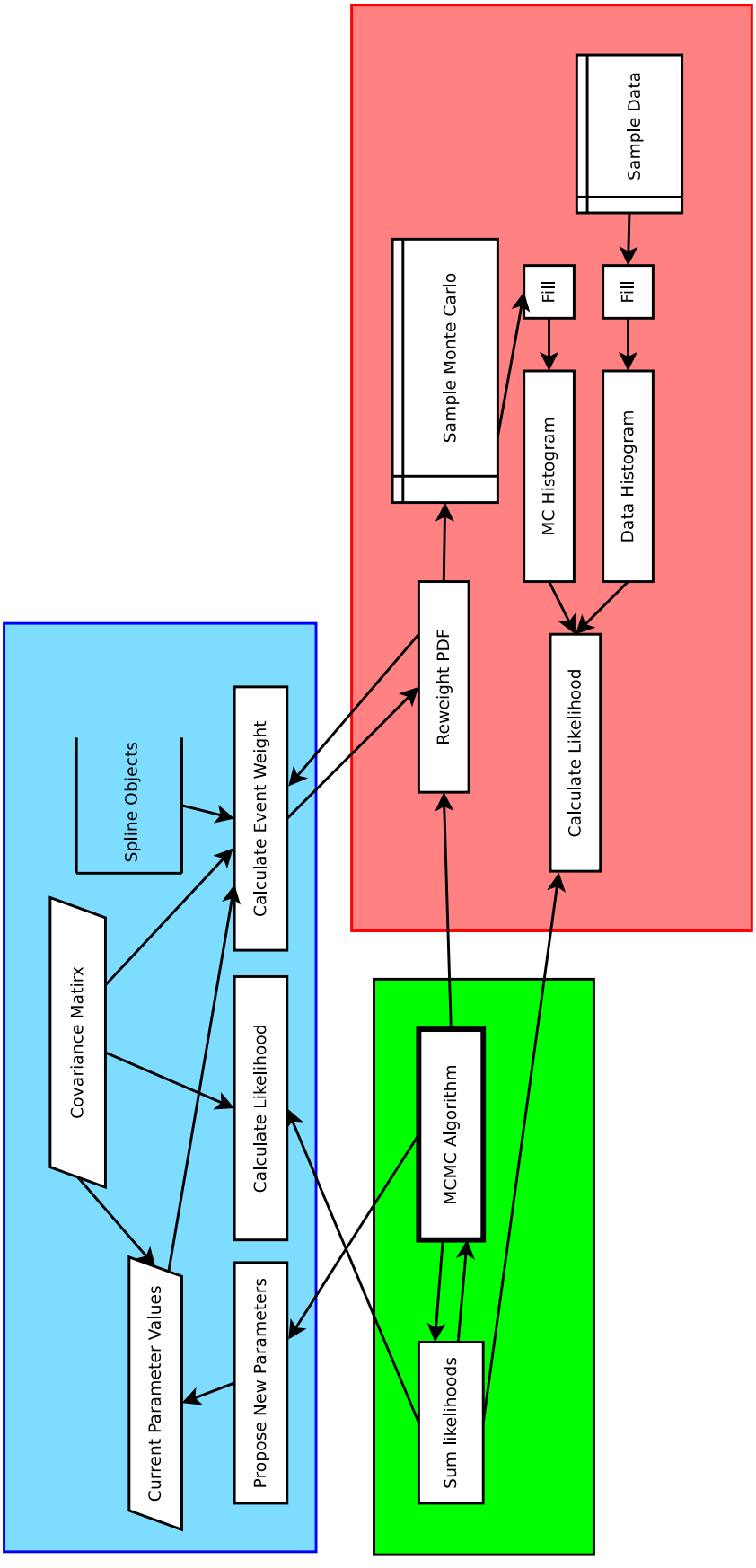


Figure 4.2.: Diagram illustrating the algorithm flow and structure of MaCh3. The green area denotes the fitting algorithm, in this case a MCMC. The blue area shows the functionality of the parameter handling classes, which handle reweighting calculations and step proposals. In red is the sample class, which handles the construction of PDFs and communicates with the parameter classes to reweight the events it contains. The modular framework allows multiple sample and parameter classes to be combined to create different analysis configurations.

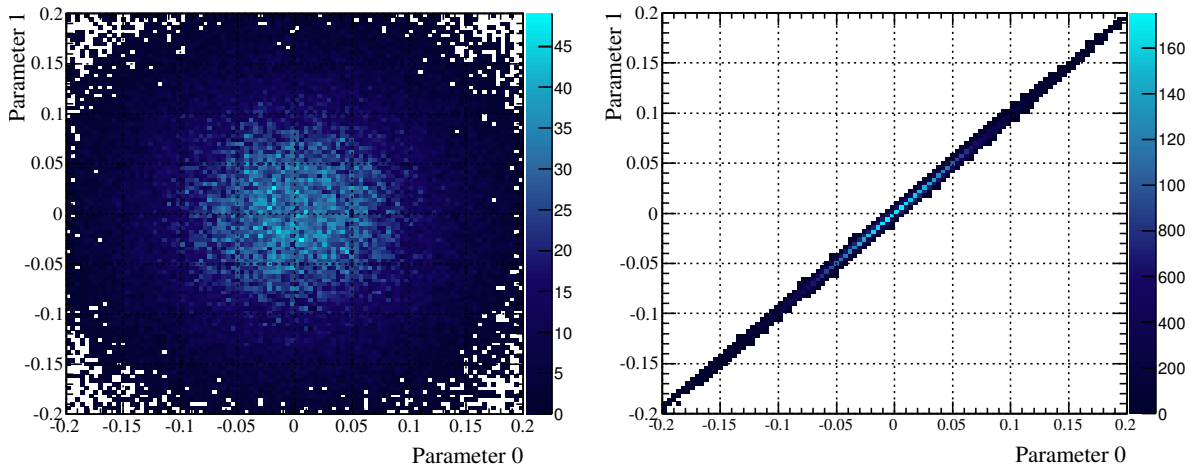


Figure 4.3.: An example of proposed steps without (left) and with (right) correlations. The distribution on the left contains many proposed steps that would not be accepted by the Metropolis-Hastings algorithm (see Section C.1), and so the efficiency of the MCMC would suffer. The correlated steps are only proposed near the region of high probability, so more steps are accepted and the MCMC will reach a stationary distribution faster.

found in Appendix C, however the mechanism for generating a new state that preserves the correlations between parameters is described here.

By using the covariance matrix used as a prior constraint on a parameter group, one can give the MCMC a good approximation of the posterior correlation between parameters. This is achieved by drawing a proposed step from a Gaussian of 1 σ standard deviation for each parameter, and then multiplying by the covariance matrix. This has the effect of imprinting the correlations and standard deviations onto the uncorrelated unit Gaussian throws, effectively generating a random draw from an N-dimensional correlated distribution.

In practise it is faster to use a decomposed covariance matrix calculated by Cholesky Factorisation [75]. This process decomposes a symmetric positive definite matrix such that

$$A = U^T U \tag{4.1}$$

where A is the original matrix, and U is the upper triangular matrix that returns the original matrix A when multiplied with its transpose. An example of correlated and uncorrelated throws can be seen in Figure 4.3.

4.1.3. Event Reweighting

To extract parameters of interest from a dataset, the model parameters must be changed to produce varying MC predictions. This could be achieved by rerunning the MC generation, however this is impractical in an analysis due to the computational costs. Instead, the response of the parameterization is encoded in covariance matrices, normalization parameters and response functions, which are used to reweight the MC to represent a change in the parameter configuration. As the MCMC proposes a new step as described in Section 4.1.2, the new set of parameters are used to reweight the detector MC to give a new MC prediction. For example, for a cross section model $\sigma(\vec{x})$, if we propose a change of parameters $\vec{x} \rightarrow \vec{x'}$, then for event i we can calculate a weight w_i

$$w_i = \frac{\sigma(\vec{x'})}{\sigma(\vec{x})} \tag{4.2}$$

where we have taken the ratio of the nominal cross section to the updated one. By applying these weights to the MC events, the response will behave in a similar way to the results of a completely reran MC. The reweighting procedure is advantageous as it drastically reduces the time taken to calculate model variations.

4.1.4. Response Functions

In T2K analyses, some parameters have a non-linear response due to strong correlations. The parameters are modelled using simulations overseen by internal working groups in T2K [77], however the code is not optimized for speed and thus is impractical to include directly in a complex analysis. Instead, the response function of each parameter is constructed using a cubic spline object. A spline is a piecewise cubic polynomial with a continuous second derivative, meaning that the transition at the boundaries between each polynomial is smooth. An example of a typical spline used can be seen in Figure 4.4.

The cubic splines are constructed by calculating the parameter response individually by scanning each parameter through variations of σ, whilst holding all other parameters

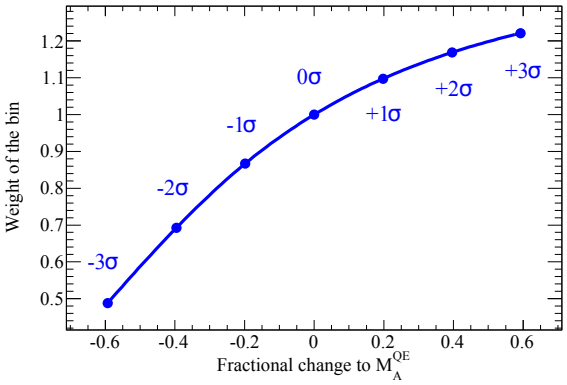


Figure 4.4.: Response function used to map the parameter response as a function of fractional change to the parameter. Figure taken from [76].

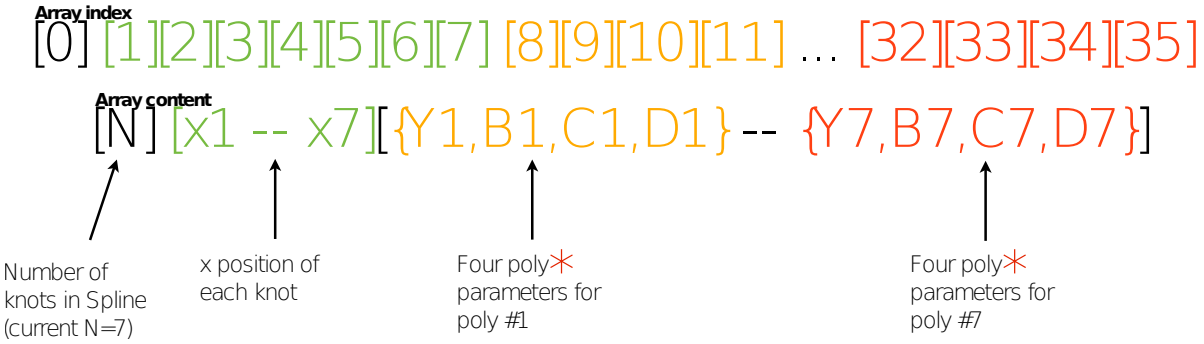


Figure 4.5.: Formatting of a single spline into an array object. Figure taken from [80].

in the model at nominal values. These data points are then turned into cubic spline objects by the `TSpline3` class from ROOT [78], with the method described in [79]. Then, for an arbitrary change to a parameter in the MaCh3 analysis framework, the fractional change is calculated by interpolating the spline.

Optimizing Spline Evaluation

MaCh3 can optimize spline evaluations by converting the `TSpline3` objects into a large array structure that is more efficient in its evaluation [80]. This structure has also been implemented on a GPU to provide an additional speed improvement.

The general idea is to deconstruct the `TSpline3` object into the basic data structures, namely an array containing the x position of the knots, and the values of the coefficients and constants for the polynomials $f(x) = ax^3 + bx^2 + cx + y$. This information is then

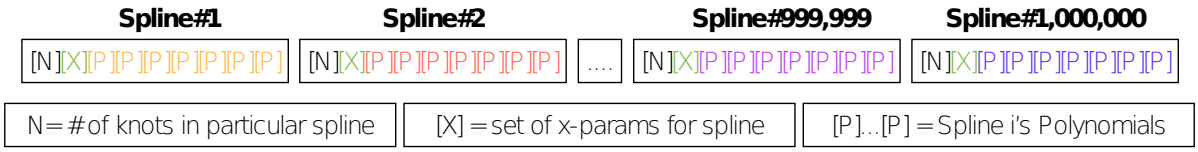


Figure 4.6.: Single array structure containing all splines in the format seen in Figure 4.5. Figure taken from [80].

structured into a single contiguous array in memory, as shown in Figure 4.5. To evaluate the spline, a single function is used `Eval(x)` where the correct polynomial to evaluate is located by sorting the x positions. Then, the polynomial parameters are extracted and the polynomial is solved.

This spline array structure is easily expanded to include any number of splines in one single contiguous memory space. As the number of array elements per spline is constant, the extended evaluation function `Eval(i,x)` can locate spline i quickly. This multi-spline structure is illustrated in Figure 4.6. The new spline structure outperforms an array of `TSpline3` objects by approximately a factor of 2 in terms of speed, and also requires less memory due to the minimalist structure and removal of memory overheads associated with C++ objects.

The multi-spline array was validated against the original splines, and an agreement to the order of 10^{-6} was found in the range of $\pm 5\sigma$, and 10^{-7} over the range of $\pm 3\sigma$ [80]. This agreement is well within the level of acceptable precision for this application.

GPU Spline Acceleration

The multi-spline format was also ported to run on a GPU using the CUDA application programming interface (see Appendix D). This involves copying the multi-spline array to GPU memory at intialization, then at every iteration of the MCMC, the new parameter fractional error changes are copied to GPU memory, and are used to evaluate all splines in parallel. The code is configured such that each spline is evaluated by a single thread on the GPU. The resulting weight for each spline is stored in a single array in GPU memory which is then copied back to the CPU RAM, where the weights are applied to events sequentially.

As the splines are constant throughout the analysis, they only need to be copied once to the GPU. This alleviates the potential latency of copying large quantities of data

from host to device, only the new parameter fractional errors need to be copied to the GPU at each MCMC iteration, and the weights copied back. This is found to be an acceptable latency which is hidden by the amount of parallel work done by the GPU; using the GPU accelerated multi-spline array method was found to have a factor ~ 20 speed improvement in the evaluation of 4 million splines.

4.2. Bayesian Parameter Inference

A Bayesian analysis aims to construct the posterior probability distribution of a hypothesis, given a data set and prior information regarding the model. This is described by Bayes' theorem:

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)} \tag{4.3}$$

where the posterior probability $p(H_i|D, I)$ is defined as the probability that the hypothesis H_i is true given the dataset D and prior information I, $p(H_i|I)$ is the prior probability, $p(D|H_i, I)$ is the probability of getting D if H_i and I are true (i.e. the likelihood function), and $p(D|I)$ is a normalization factor which ensures that the posterior probability totals to unity.

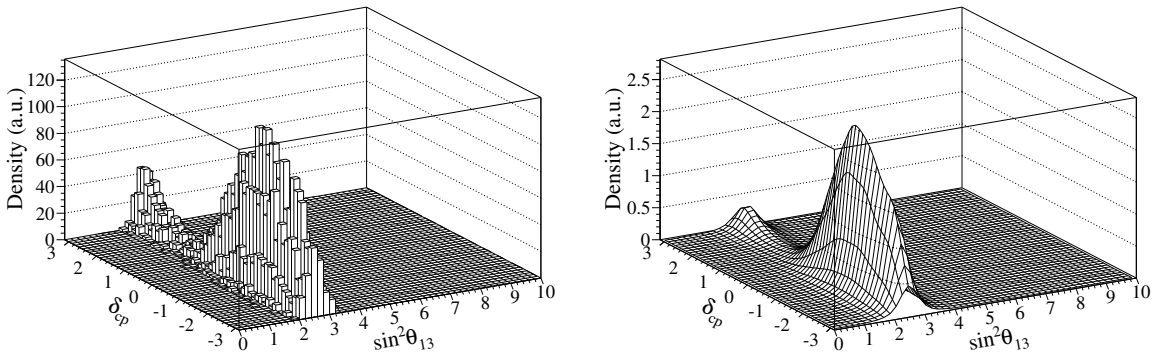
The posterior probability $p(H_i|D, I)$ gives us information about parameters, but is usually a high dimensional function, so visualizing it may not be possible, nor solving analytically. The best thing we can do is sample from it, and use this distribution of samples to approximate $p(H_i|D, I)$. As discussed at the beginning of this chapter, a MCMC is an efficient way to perform this sampling.

As physicists we endeavour to report a best-fit value and error estimation for the parameters of interest. Such information can be obtained by choosing an estimator with which to classify the posterior distribution. A simple analysis may report the mean or mode of the posterior as the best fit value, however as the number of parameters and fit complexity increases, this can become less intuitive. Typically, if the posterior has non-Gaussian features then the mean loses its credibility as a best fit value. The mode of a distribution is a better estimation of the best fit, reflecting the point of highest density and therefore most probable value of the posterior.

4.2.1. Point Estimation

Finding the mode of a multi-dimensional posterior distribution is equivalent to finding the region of highest density of a series of discrete points distributed in parameter space. One may take several approaches to estimate the point of highest density. The simplest method is to use a histogram to collate the information into binned intervals, and then the bin containing the most points is the region of highest density. This method is widely used and easy to implement, however there are several drawbacks associated with higher dimensional spaces. As the number of dimensions goes above 2 or 3, the statistics required to sufficiently populate the bins increases. A wider bin width may be used in this case, however in doing so the density estimation becomes less precise.

Kernel Density Estimation



(a) **(b)**

Figure 4.7.: Two density estimations of a 2 dimensional space. Figure 4.7a shows a histogram, and Figure 4.7b shows the KDE method. The $\sin^2\theta_{13}$ axes in both plots have been rescaled.

T2K has sensitivity to 4 oscillation parameters (θ_{23}, θ_{13}, Δm^2_{32} and δ_{cp}), and as a result, considering the posterior as a 4 dimensional histogram is impractical due to limited bin statistics. Therefore, this analysis framework uses a kernel density estimation (KDE) technique to turn a set of discrete points into a smooth continuous density function. Minuit [81] is then used to find the point of maximum density. An example of both histogram and KDE methods of density estimation is shown in Figure 4.7.

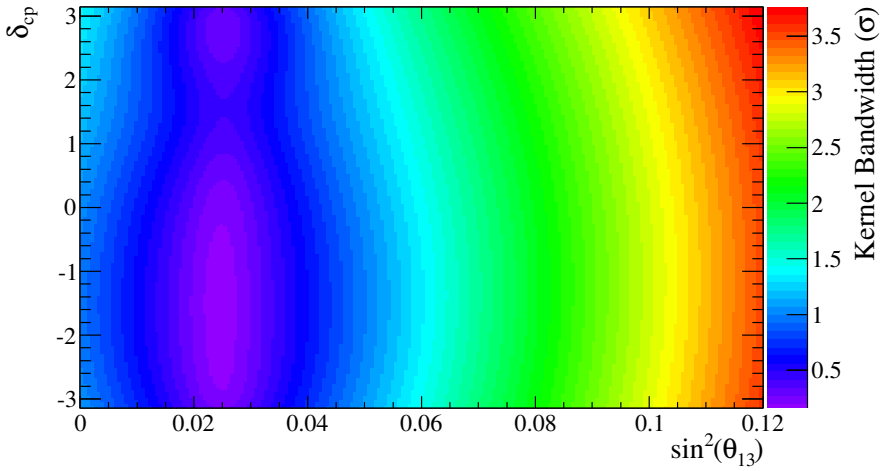


Figure 4.8.: A scan across the $\sin^2(\theta_{13})$-δ_{cp} parameter space, with $\sin^2(\theta_{23})$ and Δm^2_{32} fixed at their best fit values, and all systematic parameters marginalized. At each step, the σ of the Gaussian smoothing kernel is plotted. The normalization term of the weight for each sample is held constant. The value of the constant can be chosen, so the absolute scale of σ is arbitrary in this plot.

The kernel density estimator at a point x is defined as:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{4.4}$$

where x is the evaluation point, x_1, x_2 $\ldots x_n$ are discrete points and K is the kernel function. This framework uses a Gaussian kernel function, with bandwidth h becoming the σ of the Gaussian:

$$\hat{f}(x) = \frac{1}{n\sigma\sqrt{2\pi}} \sum_{i=1}^{n} e^{-\left(\frac{x-x_i}{\sqrt{2}\sigma}\right)^2}. \tag{4.5}$$

Adaptive Kernel Bandwidth

For optimum smoothing, an adaptive KDE is used that adjusts the bandwidth to the local density of points as detailed in [82]. In this method, the bandwidth is inversely proportional to the local density of points—producing a larger bandwidth in areas of low density and a smaller bandwidth in areas of high density—which means that low density

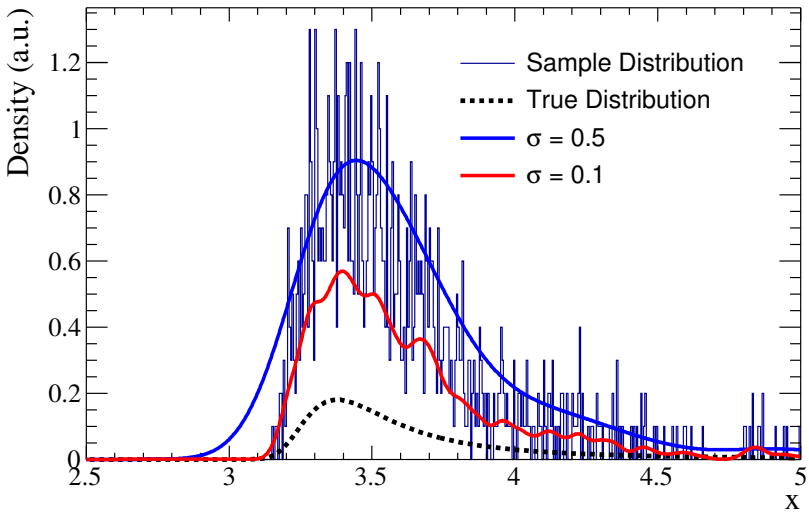


Figure 4.9.: A true Landau distribution is estimated from a density of samples using a the kernel estimation method. Shown are 2 attempts at density estimation; one where the kernel bandwidth is too large, another where the bandwidth is too small. Both methods result in the mode of the distribution being incorrectly determined. The overall normalizations of these functions are arbitrary.

areas are not undersmoothed and high density areas are not oversmoothed, as illustrated in Figure 4.9.

Adaptive kernel bandwidth is achieved by holding the normalization term $\frac{1}{n\sigma\sqrt{2\pi}}$ constant. Since n is fixed this means that the kernel width σ must be varied. In practice, the correct σ is found using a root-finding algorithm. Since the 4 parameters that make up the parameter space to be estimated have varying orders of magnitudes (e.g. $\sin^2\theta_{23} = \mathcal{O}(10^{-1})$ and $\Delta m^2_{32} = \mathcal{O}(10^{-3})$), the parameters are all scaled to constant magnitude in order to prevent issues with the bandwidth. Figure 4.8 shows the effect of this variable bandwidth as a function of the parameter space.

4.2.2. Marginalization of Nuisance Parameters

Often when considering a joint probability $p(i, j)$, we are more concerned with a single parameter, and wish to eliminate the other parameters. By calculating the marginal distribution of a parameter, one can effectively integrate over the nuisance parameters in the model. Considering the joint probability between two discrete variables x and y, one question we can ask is what is the probability that $x = i$. Using the notation

$p_x(i)$ to express this desired probability, it is easy to understand that the value of y is irrelevant to the question we are asking. Then, in discrete space we can express this marginal probability as the sum

$$p_x(i) = \sum_{j=-\infty}^{\infty} p(i,j) \tag{4.6}$$

where we have effectively summed over all possible values that the nuisance parameter y can take. Likewise, in continuous space the sum is replaced with an integral.

Consider an example where 2 dice named x and y are rolled. Table 4.1 shows the probabilities for all the combinations of throws i, j that can occur, i.e. the joint probability $p(i,j)$. If we wanted to find the probability that dice x rolls a 1, regardless of the outcome of dice y, this can be found by summing the row corresponding to that throw. This results in the marginal probability $p_x(1) = 1/6$; we have marginalized the result of the second dice. All marginal probabilities can be found by summing the corresponding row or column. In the context of an oscillation analysis, we are interested in the value of

j \ i	1	2	3	4	5	6	$p_x(i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p_y(j)$	1/6	1/6	1/6	1/6	1/6	1/6	

Table 4.1.: Joint probability distribution $p(i,j)$ of the possible throws from two dice. Information taken from [83].

a particular set oscillation parameters of interest H_{poi}. We want to marginalize out all other parameters so we can report the most probable values and uncertainties of H_{poi} taking into account all uncertainty introduced by the nuisance parameters. Following the logic of the previous example, to construct the marginal posterior for our parameters of interest $p(H_{poi}|D,I)$ (i.e. create a projection of a subset of parameters in a smaller number of dimensions) we want to integrate over all other potential combinations of

nuisance parameters

$$p(H_{poi}|D,I) = \frac{\int d\theta p(H_{poi},\theta|I)p(D|H,\theta,I)}{p(D|I)}, \tag{4.7}$$

where θ are the nuisance parameters, D is the data, I is prior information and H is all hypotheses, in this case we mean all parameters including our parameters of interest. In practise this integral is non-trivial, as each parameter requires its own integral to be introduced.

4.2.3. Credible Intervals

From the Bayesian posterior one can extract an estimate of the posterior uncertainty by looking at the distribution of probability. One such way to do this is by constructing Bayesian credible intervals. Credible intervals contain a fraction α of the total probability in the posterior density. For example, the integral

$$\int \pi(\theta|x)d\theta = 1-\alpha, \tag{4.8}$$

will contain $(1-\alpha)100\%$ of the probability of the posterior distribution.

MaCh3 uses highest posterior density (HPD) credible intervals, which are defined such that any point within the interval is more probable than outside [84]. This results in a definition that aims to construct the smallest credible region in 1-dimension, which is suitable for reporting parameter sensitivities as is required of an oscillation analysis. As all reported credible regions are presented in a maximum of 2 dimensions, credible intervals are constructed using a 1- or 2-dimension histogram. The histogram bins are ordered from highest populated to lowest, and then the bins are counted from most to least dense until the bins counted so far contain the required fraction of the total probability of the posterior. For example, to report 68% credible regions, the bins are counted until 68% of the posterior integral is contained within the counted bins.

4.2.4. Model Comparison

Often it is desirable to test various models on the same data set. Bayesian statistics allows the comparison of models in a similar way to that of parameter estimation. In the same way a parameter's posterior PDF is proportional to the prior multiplied by

the likelihood, the posterior probability for the whole model is proportional to the prior probability multiplied by the global likelihood. Using Bayes' theorem in Equation 4.3 we can set $H = M$ see the probability for a model M to be true given a data set D. Using 2 models labelled i and j, and assuming as prior information that one of them is true, it is useful to consider the odds ratio:

$$O_{ij} = \frac{p(M_i|D,I)}{p(M_j|D,I)} \tag{4.9}$$

where we have computed the ratio of M_i being favoured over M_j. Now inserting Equation 4.3 into 4.9, we see

$$O_{ij} = \frac{p(M_i|I)p(D|M_i,I)}{p(M_j|I)p(D|M_j,I)} \equiv \frac{p(M_i|I)}{p(M_j|I)} B_{ij} \tag{4.10}$$

where B_{ij} is the Bayes factor, which can be used as an estimate of how favoured the model M_i is compared to M_j [85]. The strength of evidence that model M_i is favoured over M_j can be judged from Table 4.2.

Bayes factor	Strength of Evidence
$-\infty < B_{ij} \leq 0.1$	Strongly against M_i
$0.1 < B_{ij} \leq \frac{1}{3}$	Substantial against M_i
$\frac{1}{3} < B_{ij} < 1$	Weak against M_i
$1 \leq B_{ij} < 3$	Weak in favour of M_i
$3 \leq B_{ij} \ll 10$	Substantially in favour of M_i
$10 \leq B_{ij} < \infty$	Strongly in favour of M_i

Table 4.2.: Interpretation of the Bayes factor B_{ij}. Information taken from [85].

4.3. Framework Summary

The MaCh3 analysis framework provides a set of base classes from which samples and models of systematic uncertainty can be constructed. These classes can be combined in different configurations, in order to produce different analyses of data sets. At the heart of the framework is the MCMC algorithm class, which constructs the Bayesian posterior probability distribution for the model constructed by the analyser. All information about

the parameters of interest is inferred from this posterior, and so the framework also contains the tools to effectively extract marginalized credible intervals and most probable values in multiple dimensions, using density estimation techniques.

Though it is not a requirement, MaCh3 is capable of event-by-event reweighting techniques, whereby all PDFs are constructed by reweighting individual weights, instead of using templates. This enables the optimal use of the information contained within the detector MC. The use of event-by-event reweighting comes at a cost. The time taken to compute a single step in the MCMC is increased by orders of magnitude, and so to make this technique feasible, computing intensive algorithms are off-loaded onto GPUs where the execution time is drastically reduced. Chapter 5 details the study and validation of calculating oscillation probability on an event-by-event basis on GPUs. Section 4.1.4 described the acceleration of spline evaluations on GPUs.

Chapter 5.

Parallel Event-by-Event Reweighting Calculations on GPUs

5.1. Introduction

To construct the MC prediction necessary for an oscillation analysis, the probability distribution function (PDF) must be constructed empirically from detector MC and reweighted according to both the neutrino oscillation model and any corresponding systematic uncertainties. The analysis presented in Chapter 7 uses the standard three flavour neutrino oscillation model, with the PMNS transition matrix describing the strength of the mixing, as defined in Equation 1.2. The probability that a MC event will be observed either as the same or a different flavour is calculated using the true neutrino energy (see Equation 1.11), along with parameters such as baseline and earth density. This is illustrated in Figure 5.1.

The presence of matter induces non-negligible effects on ν_e and $\bar{\nu}_e$ due to forward scattering on electrons in matter as discussed in Section 1.2.3. These so-called matter effects add computational complexity but can be calculated as prescribed in [10].

Event-By-Event Reweighting

Neutrino oscillation analyses are often performed by producing a large sample of simulated events in order to estimate the PDF, as many reconstruction effects may be complex. These simulated events are produced at a certain set of oscillation parameters and

Table 5.1.: Assumed oscillation parameters for all studies presented.

Parameter	Value
$\sin^2(\theta_{12})$	0.311
$\sin^2(\theta_{23})$	0.5
$\sin^2(\theta_{13})$	0.0251
Δm^2_{32} (eV^2)	2.4×10^{-3}
Δm^2_{12} (eV^2)	7.6×10^{-5}
δ_{cp}	0
Earth Density (g/cm^3)	2.6
Baseline (km)	295

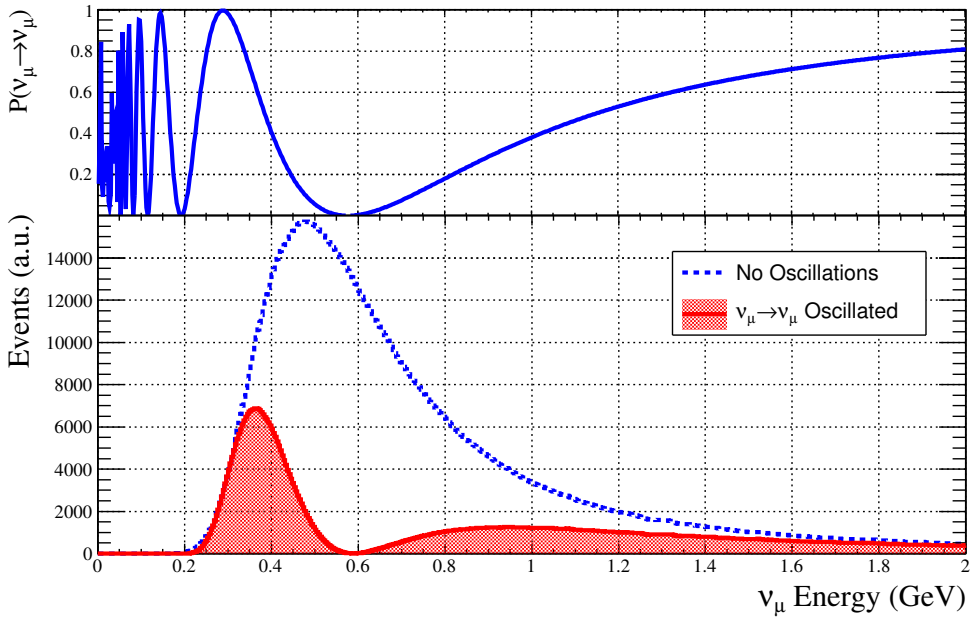


Figure 5.1.: *Top*: $\nu_\mu \rightarrow \nu_\mu$ neutrino survival probability calculated with matter effects for a propagation distance of 295 km through a constant matter density of 2.6 g/cm^3 . *Bottom*: A mock ν_μ neutrino beam spectra under the influence of this oscillation probability, compared to the no oscillation case. The trough of the oscillation probability function can been seen to line up with the trough of the oscillated spectra at 0.6 GeV. Oscillations were calculated using parameter values listed in Table 5.1 with normal hierarchy.

experimental parameters, all of which must be varied in order to find the optimal output parameters for analysis. Binned maximum likelihood analyses are an effective way to compare the data to the MC to optimize the parameters. Calculating the effect of the variation of oscillation and systematic parameters can be done in two ways for these binned MC PDFs. One option is to calculate the effect of the variation at the centre of each bin and apply it to the whole bin; this has the advantage of being relatively quick, but the disadvantage of losing any shape information which resides within the bin boundaries. The other option is to retain all of the simulated events and calculate the variations on an event-by-event basis; this has the advantage of retaining any shape information within the bin, but the disadvantage of requiring many more calculations.

Both oscillation parameters and systematic uncertainty parameters are subject to this binning effect. An example of a systematic uncertainty that would be affected by binning is a scale uncertainty for energy reconstruction, critical for oscillation analyses. Using a binned weighting method loses the information about the reconstructed energy of any given event, and so produce a different predicted number of events than simply scaling the true reconstructed energy of the constituent MC events. Further discussion of systematic uncertainties is beyond the scope of this chapter, but it comprises part of the motivation to find a computationally efficient way to treat the constituent MC events individually.

The binning effect on oscillation parameters can be as large as a few percent. One can see this effect by placing a histogram bin with a typical width of 25 MeV from 0.6 GeV to 0.625 GeV (near to the oscillation maximum shown in Figure 5.1). Considering the case of integrating the true neutrino energy spectrum in this bin and multiplying by the oscillation probability at the bin centre (0.6125 MeV), and comparing this with the result of integrating the product of the oscillation probability and the input neutrino spectrum one finds a difference of 2.6%. This difference arises from the approximation that all neutrinos within the bin edges have the same true energy.

This is a strong motivation to find a way to treat the constituent MC events according to their true properties. This method increases the number of oscillation weight calculations by several orders of magnitude, and thus can sometimes inhibit the use of such a computationally expensive technique on a CPU. This chapter details the use of a graphics processing unit (GPU) to perform this event-by-event reweighting in parallel. An explanation of GPUs and their use for non-graphical purposes can be found in Appendix D.

5.2. Implementation on a GPU

A typical CPU consists of ~ 4 cores with clock speeds in the range of 3-4 GHz and has the capacity to run multi-threaded applications. In contrast, a modern consumer GPU has 100-1000 cores running at ~ 1 GHz that are used for graphical calculations, however the architecture can now be exposed for non-graphical applications with application programming interfaces (API) such as the *Compute Unified Device Architecture* (CUDA) [86] and *Open Computing Language* (OpenCL) [87]. Such *General Purpose Graphics Processing Units* (GPGPU) can greatly outperform a CPU if a problem can be parallelized accordingly.

Because each event in a Monte Carlo sample is independent, oscillation weight calculations can be performed in parallel. The library `Prob3++` [88] was ported to the GPU using the CUDA API to enable fine-grained concurrent calculations. The results displayed in Figure 5.2 show the execution times for varying numbers of calculations in series (CPU) and parallel (GPU). Also compared is the original code using the *Open Multi-Processing* (OpenMP) API, which allows the code to use multiple CPU cores [89].

5.2.1. Method

A series of C/C++ algorithms for calculating oscillation probabilities were ported to CUDA. Functions that execute on the device must be compiled separately by the `nvcc` compiler provided by NVIDIA and linked into the host program using a compiler such as `gcc` [90].

Within the GPU code, an array of energy values were allocated and instantiated in host memory (the system's RAM) and then copied to the device memory (the graphics card's video RAM) using API function calls provided by CUDA.

In addition to the event energies, components that are dependent only on the oscillation parameters (i.e. Equation 10 of [10]) are computed on the CPU and then copied to the GPU in the same manner as the energy array.

The calculations in `Prob3++` were modified into a set of CUDA kernel functions (functions that run in parallel on the GPU) and were then executed on each element of the array in parallel, which performs the oscillation probability calculation in double precision. The result of this calculation is written to an array in the device memory, and

is then copied back to the host. All memory allocation and transfer operations to and from the GPU device are handled via CUDA API functions. A simplified example of this process can be found in Listing 5.1.

Listing 5.1: Example of copying data to GPU memory and executing a kernel.

```
// size of array
size_t size = n * sizeof(double);

// allocate host memory
double *true_energy_host = (double*) malloc(size);
double *osc_weight_host = (double*) malloc(size);

// allocate device memory
double *true_energy_dev = cudaMalloc((void **) &true_energy_device, size);
double *osc_weight_dev = cudaMalloc((void **) &osc_weight_device, size);

// fill energy array
...

// copy energy array to the device
cudaMemcpy(true_energy_dev, true_energy_host, size, cudaMemcpyHostToDevice);

// instantiate and perform copy of mixing matrix
...

// execute GPU kernel on the array
calculateOscProb<<<gridsize, blocksize>>>(...);

// copy the results back to the host
cudaMemcpy(osc_weight_host, osc_weight_dev , size, cudaMemcpyDeviceToHost);
```

5.2.2. Results and Validation

The Comparison of CPU vs. GPU execution times as a function of number of events reweighted shows the CPU performing better at small number of events, with the GPU performing up to 132 times faster at 1.45 million calculations (Figure 5.2). The “crossover” point is hardware dependent, and is expected to change with different CPU/GPU

combinations, and also different algorithm implementations. At best, the multi-threaded code gains only 2-3 times speed improvement. Figure 5.3 shows the speedup relative to single core execution. As seen in Figure 5.2, the execution time of the GPU implementation remains steady until it reaches a point where all threads are occupied and the limit of concurrent execution is reached [91].

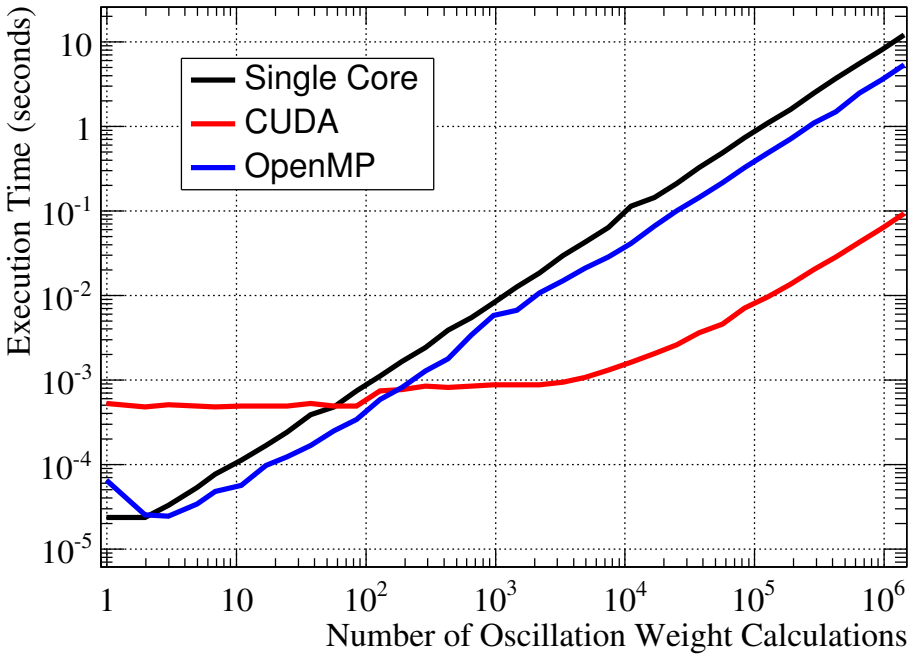


Figure 5.2.: Comparison of execution time for varying numbers of calculations between CPU and GPU implementations. The plateau observed in the CUDA results is due to the total number of threads not yet fully occupied. At 10^3-10^4 number of calculations, the GPU becomes saturated and starts to execute in series.

The overheads associated with copying to and from host and device memory across the PCI-E bus can be a large source of latency, and as can be seen in Figure 5.2, the CPU will outperform the GPU if the number of concurrent calculations is small.

To validate the GPU code, 10 million random energy values were drawn from a uniform distribution between 0 and 30 GeV, and were used to calculate oscillation weights on CPU and GPU. The residuals between CPU and GPU calculations were found to be on the order of 10^{-12} for double precision, and are plotted in Figure 5.4. The residual is attributed to the difference between hardware implementations of arithmetic operations [92], and in this test is considered negligible.

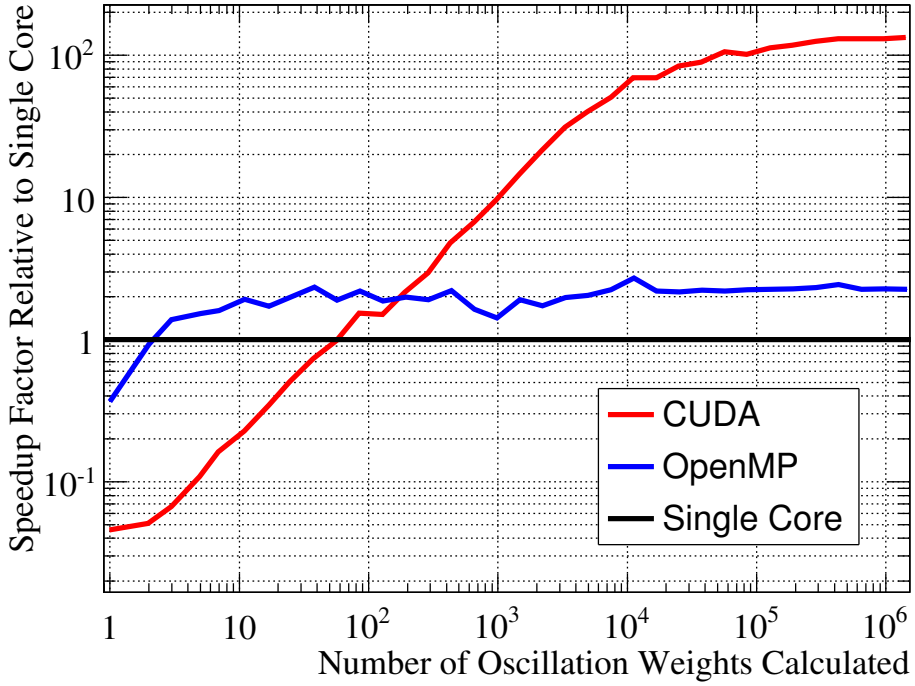


Figure 5.3.: Speedup relative to single core execution.

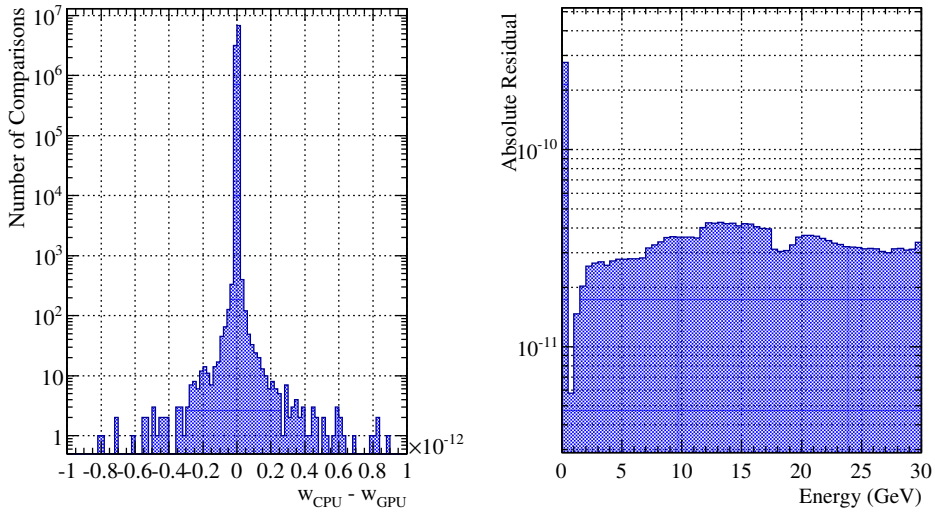


Figure 5.4.: *Left:* Residuals between weights calculated on CPU w_{CPU} and GPU w_{GPU} for the same oscillation parameters and value of energy. *Right:* The absolute difference between energy spectra weighted by w_{CPU} and w_{GPU}.

The GPU implementation and original version of `Prob3++` were also compared within a simple toy oscillation fitter written using the *Bayesian Analysis Toolkit* [93]. The motivation is to give a realistic measure of speed improvement for an application in a physics analysis, as well as to show that there is negligible difference between both CPU and GPU methods when used in a realistic way. The fit uses a Markov chain Monte Carlo to sample the oscillation parameter space, building a Bayesian posterior density via the Metropolis-Hastings algorithm, from which credible intervals can be constructed. The likelihood function is defined as:

$$L(\vec{o}, \vec{f}|\vec{D}) = \prod_i p(\vec{D}|\vec{o}, \vec{f}) \tag{5.1}$$

where $\vec{o}$ are the two parameters of interest θ_{23} and Δm^2_{32}, $\vec{f}$ are the nuisance parameters $\theta_{12}, \theta_{13}, \Delta m^2_{12}$ and δ_{cp}, and p is the probability mass function of a dataset $\vec{D}$ given parameters $\vec{o}$ and $\vec{f}$. The toy fit simulates a long baseline ν_μ disappearance analysis by fitting a fake ν_μ far-detector energy spectra $\vec{D}$, created by sampling from a Landau function and weighted using the oscillation parameters found in Table 5.1.

The PDF is constructed by taking a large number of samples (on the order of millions) from the Landau distribution and binning these samples into a histogram weighted by the oscillation probability calculated with `Prob3++`. An example of oscillated and unoscillated spectra can be seen in Figure 5.1.

As the Markov chain Monte Carlo proposes a new set of oscillation parameters each step, the PDF is reconstructed using the event-by-event method described above and compared to the data. Therefore the calculation of oscillation weights provides a large overhead to the fit method and is directly related to the calculation of likelihood.

Out of five oscillation parameters, all but θ_{23} and Δm^2_{32} are fixed at the values listed in Table 5.1. θ_{23} and Δm^2_{32} are free to float in the fit, and have flat prior distributions meaning that there are no prior constraints on these parameters in the likelihood function. This constructs a 2D posterior distribution in θ_{23}-Δm^2_{32} space.

The best fit and error value of the fitter was compared between CPU and GPU oscillation reweighting methods. The difference between CPU- and GPU-made spectra and posterior distributions using identical oscillation parameters was found to be to an acceptable precision, and plotted in figure 5.5. Furthermore, an order of magnitude speed increase was observed for the overall fitting procedure by off-loading oscillation reweighting to the GPU.

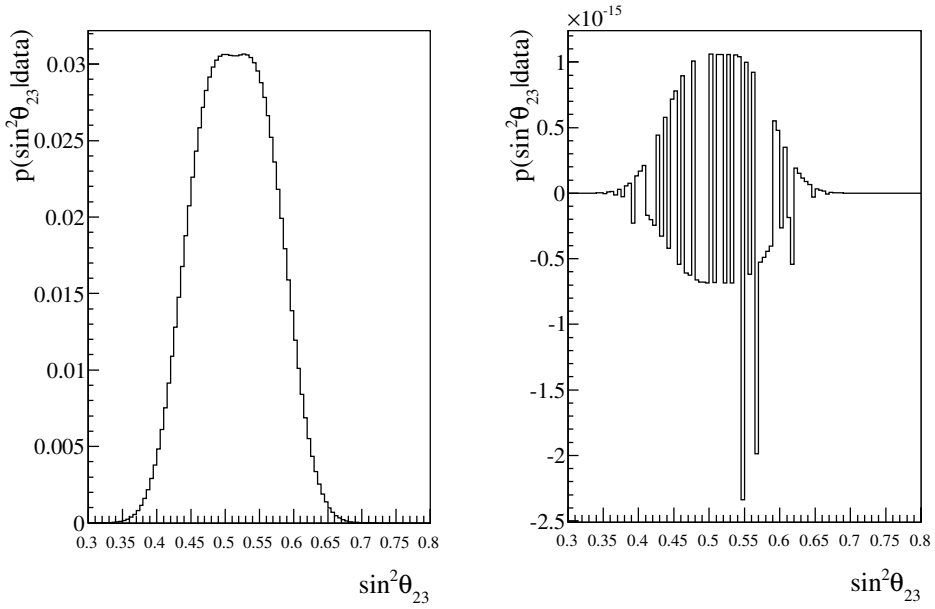


Figure 5.5.: *Left:* 1-dimensional $\sin^2(\theta_{23})$ marginal distribution. *Right:* Difference between the 1-dimensional marginal distribution of $\sin^2(\theta_{23})$ generated on CPU and GPU. The marginal distribution encodes information about the most probable value and the uncertainty of the parameter.

The results presented are prepared using an Intel Xeon E5640 quad-core processor running at 2.67 GHz, and an NVIDIA M2070 GPU with 448 CUDA cores running at 1.15 GHz. The code is compiled for 64-bit hardware using the gcc compiler version 4.6.3 with the -O2 optimization flag, and the CUDA toolkit version 5. OpenMP code is restricted to use 4 threads which ensures execution on the physical cores of the CPU.

5.3. Summary

The parallel implementation of oscillation reweighting enables the improvement of neutrino analyses via the computation of Monte Carlo weights on an event-by-event basis, which is a limiting factor of an analysis if performed solely on a CPU. The analysis presented in Chapter 7 makes use of this GPU reweighting to make event-by-event reweighting feasible. Event-by-event reweighting retains all the Monte Carlo spectral shape information that is otherwise lost when binned into a histogram. More importantly, by being able to discriminate events within a sample of Monte Carlo, event migrations can be modelled, and as T2K collects more statistics this systematic effect will become more prominent.

5.4. Impact of GPU Technology

The oscillation analysis presented in Chapter 7 made extensive use of GPU technology to make feasible the event-by-event reweighting method. This not only enabled a better treatment of systematic uncertainties, and more accurate representation of the oscillation probabilities effect on the MC predictions, but also has financial and environmental benefits pertaining to the operational costs and carbon footprint of the hardware used to perform the analysis.

A typical analysis makes use of a computing cluster, and this analysis is no different. In fact, it uses the EMERALD GPU cluster to enable work to be distributed across many nodes, each with a GPU to perform individual reweighting. At first glance, one may argue that the financial burden of using a GPU cluster is higher, because of the increased hardware and electricity costs. However, with a quick "back of the envelope" calculation we see that it is in fact more cost effective to use GPUs.

Assuming the electricity rate of 1 kWh = £0.15, and that all hardware runs at its maximum total dissipated power (TDP), a basic cost analysis can be performed.

If the analysis uses 100 computing nodes on the Emerald cluster to perform a complete analysis, then using only CPU, the analysis takes approximately $11 \times 20 = 220$ h. The Xeon X5850 processors have a maximum TDP of 95 W, so this results in a total power consumption of 2.09 MW, costing £313.50. In Contrast, when running in GPU mode, we have the added TDP of the M2090 GPU, which totals 95 + 225 W. However, with the speed up factor of 20, the analysis only takes 11 h to complete and thus uses 0.352 MW, costing £52.80. We now see that although the initial overhead of buying GPU hardware is high, the money saved in running costs is significant.

Chapter 6.

MC Prediction of SK and ND280 Samples for a 3 Flavour Joint Oscillation Analysis

T2K was designed to make precision measurements of the already known neutrino mixing parameters, and search for non-zero θ_{13}. T2K's beam is composed mainly of ν_μ, and so there are two ways to observe neutrino oscillation at the far detector; by looking for a deficit of ν_μ, and looking for an appearance of ν_e. Selections of neutrino events at SK are made by choosing either muon- or electron-like candidates by discriminating between the different ring patterns produced by charged particles. Both of these selections probe different combinations of mixing parameters, and so it makes sense to analyse them together. The combination of both ν_μ disappearance and ν_e appearance samples is referred to as a *joint* analysis. The analysis performed here involves constructing a likelihood function using contributions from detector data, and prior information about the parameterization of the oscillation and systematic model.

To constrain dominant beam flux and neutrino cross section uncertainties at SK, three muon event topologies were selected at ND280 and analysed simultaneously with the SK samples. This is achieved by adding an extra term to the likelihood function, involving the ND280 data and relevant systematic parameters. This is in contrast to previous methods used for T2K oscillation analysis publications, where the result of the ND280 analysis was propagated to the oscillation analyses via a multivariate Gaussian constraint of the correlated systematic parameters; no ND280 data term was used.

This chapter details the empirical construction of MC predictions from both ND280 and SK detector samples. These MC predictions will be used to calculate the likelihood term for the data samples from all detectors. Chapter 7.2 details the validation and results of the analysis of the first 4 T2K data runs (Run 1-4) using SK ν_μ disappearance and ν_e appearance samples to extract the atmospheric mixing parameters θ_{23} and Δm^2_{32}, whilst simultaneously fitting ND280 near detector ν_μ samples to constrain the dominant systematics. The combination of disappearance and appearance samples introduces sensitivity to θ_{13}. By exploiting recent constraints on θ_{13} from reactor experiments [94], we will see that a constraint on the δ_{cp} phase emerges from the tension between the data.

The analysis is written using the MaCh3 framework described in Chapter 4. The analysis is validated via statistical methods.

6.1. Introduction

This analysis takes a Bayesian approach to parameter inference by using a MCMC to marginalize all systematic parameters. A detailed description of MCMC can be found in Appendix C. Marginalization was explained in Section 4.2.2, and essentially means that we can express our parameter values and errors independently of what the true value of the systematic parameters are, especially in the case where the uncertainty is non-Gaussian in nature. Bayesian parameter inference uses Bayes' Theorem (Equation 4.3) to construct a posterior probability distribution, by combining the likelihood function of a model with prior information. In the case of this analysis, the likelihood function is constructed by comparing the data to an MC prediction. Prior information enters in in the form of multivariate Gaussian terms that come from various systematic sources. The MCMC is an efficient algorithm for producing a distribution of points in high dimensional space, that can be used to approximate the posterior distribution by estimating the density. As this analysis has a high number of parameters, the MCMC technique is particularly efficient when compared to other methods to perform high dimensional integrals.

Once the posterior probability distribution is obtained, one can make inferences from it by looking at the region of highest density. The most likely value of parameters along with their uncertainties are extracted using methods detailed in Chapter 4.

A binned maximum likelihood method is used, however, motivated to utilize maximum information from the detector MC, all probability distribution functions (PDFs) are

constructed event-by-event. This means that for each iteration of the MCMC, each detector MC event is reweighted and then filled into a histogram (constructing the MC prediction), which is used in the binned likelihood calculation. This is in contrast to the "template" method of filling a histogram once with the detector MC, and then reweighting the histogram bin centres to make the MC prediction. This means that the shape information within the bin is retained as an event near the bin edge may have a true energy significantly different from the bin centre, and also enables the modelling of systematic event migrations between bins and samples. If the template method is used, individual event information is lost and it is difficult to simulate events moving across a bin boundary. The oscillation probability calculations for each event are calculated on a graphics processing unit (GPU) as described in Chapter 5.

6.2. Likelihood Definition and Choice of Prior Distributions

The Metropolis-Hastings algorithm (see Section C.1) is the critical part of the MCMC that uses the likelihood function to decide whether to transition to a proposed step in parameter space. At any point in parameter space, a new state is proposed, and the likelihood of that state is compared to the current. If the proposed state is more probable, then the MCMC is more likely to step there. The MCMC does not always step to more probable states, as this would simply make the MCMC walk to the maximum of the likelihood, which would not allow proper error estimation. It is this mechanism that creates the distribution of points that will be proportional to the posterior probability distribution.

This analysis uses a likelihood-ratio function that is the sum of the probability mass function for the data samples from both detectors, and Gaussian priors for systematic parameters constrained by a covariance matrix as seen in Equation 6.1. Essentially, by sampling from Equation 6.1, we can construct a collection of points that are proportional to the posterior probability defined by Bayes' theorem. As we are free to normalize our collection of points to unity, the denominator in Bayes' theorem can be neglected.

$$
\begin{aligned}
-\ln(L) = & \sum_{i}^{ND280bins} N_i^p(\vec{b},\vec{x},\vec{f},\vec{d}) - N_i^d + N_i^d ln[N_i^d/N_i^p(\vec{b},\vec{x},\vec{f},\vec{d})] \\
& + \sum_{i}^{SK1R_\mu bins} N_i^p(\vec{b},\vec{x},s\vec{k}d) - N_i^d + N_i^d ln[N_i^d/N_i^p(\vec{b},\vec{x},s\vec{k}d)] \\
& + \sum_{i}^{SK1R_e bins} N_i^p(\vec{b},\vec{x},s\vec{k}d) - N_i^d + N_i^d ln[N_i^d/N_i^p(\vec{b},\vec{x},s\vec{k}d)] \\
& + \frac{1}{2}\sum_{i}^{E_\nu bins}\sum_{j}^{E_\nu bins} \Delta b_i (V_b^{-1})_{i,j} \Delta b_j \\
& + \frac{1}{2}\sum_{i}^{xsecpars}\sum_{j}^{xsecpars} \Delta x_i (V_x^{-1})_{i,j} \Delta x_j \\
& + \frac{1}{2}\sum_{i}^{fsipars}\sum_{j}^{fsipars} \Delta f_i (V_f^{-1})_{i,j} \Delta f_j \\
& + \frac{1}{2}\sum_{i}^{nd280det}\sum_{j}^{nd280det} \Delta d_i (V_d^{-1})_{i,j} \Delta d_j \\
& + \frac{1}{2}\sum_{i}^{skdet}\sum_{j}^{skdet} \Delta skd_i (V_{skd}^{-1})_{i,j} \Delta skd_j
\end{aligned}
\tag{6.1}
$$

The components of Equation 6.1 are as follows: V_{ij} are the elements of the covariance matrices that constrain the systematic parameters that have a Gaussian prior distribution. These are labelled b for flux, x for cross section, f for final state interactions (FSI), d for ND280 detector, and skd for SK detector. N_i^p is the number of predicted events per bin, given the values of the systematic parameters, and N_i^d is the number of events in the data sample. The systematic parameters will be discussed in more detail in Section 6.4.

PDFs and data are binned into 2D lepton momentum-$\cos\theta$ space for ND280 data and 1D reconstructed energy space for the SK samples. 2D Momentum-$\cos\theta$ PDFs are not used for SK because with current statistics there is not much additional sensitivity to be gained from the extra shape information. It is expected that future analyses will incorporate 2D PDFs for SK.

The parameters of interest are θ_{23}, θ_{13}, δ_{CP}, and $|\Delta m^2_{32}|$, and are given flat prior distributions with boundaries $0 \leq \sin^2\theta_{23} \leq 1$, $0 \leq \sin^2\theta_{13} \leq 1$, $-\pi \leq \delta_{CP} < \pi$, and $1\times10^{-3}\text{eV}^2 \leq |\Delta m^2_{32}| \leq 4\times10^{-3}$. A flat prior distribution means that there is essentially

no likelihood term for that parameter, as it assumes that *a priori* all values of that parameter are equally probable. However, to enforce the boundary conditions, if the MCMC proposes a step outside of the allowed range of the parameter, the proposed state receives a large likelihood penalty which makes the probability of that state extremely unlikely.

Additionally, the mass hierarchy (MH) is treated as a discrete variable and has a prior with equal probability in each mass hierarchy state, i.e. normal hierarchy (NH) and inverted hierarchy (IH). This corresponds to P(NH)=P(IH)=0.5 and is in practice encoded as the sign of Δm^2_{32}. This means that when the MCMC proposes a new step, there is 50% chance that the proposed value of Δm^2_{32} will change sign (and therefore change between NH and IH). By allowing the MCMC to step in both NH and IH states with equal probability, the posterior distribution contains information about which MH is favoured by the data, if any.

Systematic parameters have Gaussian priors as defined by their covariance matrices, with exception to the spectral function parameters, which have flat priors bound between 0 and 1. The spectral function flat prior boundaries are enforced in the same way as for the oscillation parameters.

6.3. Event Selection

Selections from MC to create MC predictions for SK and ND280 are discussed in this section. The total protons on target (POT) collected by each detector is tabulated in Table 6.1. The difference in total POT collected by each detector is due to different detector downtimes. This analysis uses SK MC version 13a, and ND280 MC production 5F.

6.3.1. ND280 Tracker ν_μ

ND280 is used to constrain the flux and cross-section parameters correlated between near and far detectors. This is achieved by selecting muon samples from the data in order to create a charged-current inclusive (CCInc) sample that is divided into three subsamples: charged-current 0-π (CC0π), charged-current single π^+ (CC1π), and charged-current other (CCoth). The sample is subdivided in order to isolate topologies

Run Period	Dates	Integrated POT by ND280	Integrated POT by SK
Run 1	Jan. 2010-Jun. 2010	0.17×10^{20}	0.32×10^{20}
Run 2	Nov. 2010-Mar. 2011	0.78×10^{20}	1.11×10^{20}
Run 3	Mar. 2012-Jun. 2012	1.56×10^{20}	1.58×10^{20}
Run 4	Mar. 2013-May 2013	3.39×10^{20}	3.56×10^{20}
Total	Jan, 2010-May 2013	5.90×10^{20}	6.57×10^{20}

Table 6.1.: Details of POT accumulated by the near and far detectors.

of interest for constraining cross-section systematics; namely CC0π attempts to select CCQE interactions, CC1π selects resonant pion production and CC other targets DIS. These modes were discussed in Section 2.1.

The inclusive sample (i.e. identifying a selection containing all three subsamples) is compiled using the following cut selection:

1. Good Data Quality: the global ND280 data quality flag must be good
2. Bunching: Tracks considered part of the same event must be in the same beam bunch
3. TPC Quality and Fiducial Volume: There must be at least one track beginning in FGD1's fiducial volume, and entering a TPC with at least 18 vertical TPC clusters
4. Backwards-going and TPC1 veto: if there is activity in TPC1, or if the end position of the highest momentum track is more upstream than the start position, the track is vetoed
5. Broken Tracks FGD1: Events are rejected when the muon candidate's z start position is more than 425 mm away from the FGD1 upstream edge and in the same event where at least one "FGD-only" track with its start position out the FGD1 fiducial volume exists.
6. Muon PID: The highest momentum negative track in the event must be muon-like, according to TPC PID

From this selection, the three subselections are formed.

The CC0π sample is defined by rejecting events with any pion reconstructed in the TPC, any electrons or positrons in the TPC, or any Michel electrons or pions reconstructed in the FGDs.

The CC1π sample is defined by rejecting events with negative pions or electrons or positrons in the TPC and selecting events where there is one reconstructed positive pion or one Michel electron reconstructed in the TPCs and FGDs.

The CC other sample contains all other CC inclusive events not selected by the CC0π or CC1π samples.

The binning for the samples is chosen to be as fine as possible while still requiring at least 25 MC events in each bin to reduce statistical uncertainty. The binning schemes are:

- CC0π and CCoth
  - p_μ (MeV): 0, 300, 400, 500, 600, 700, 800, 900, 1000, 1250, 1500, 2000, 3000, 5000, 30000
  - $\cos\theta$: -1.0, 0.6, 0.7, 0.8, 0.85, 0.9, 0.92, 0.94, 0.96, 0.98, 0.99, 1.0
- CC1π
  - p_μ (MeV): 0, 300, 400, 500, 600, 700, 800, 900, 1000, 1250, 1500, 2000, 5000, 30000
  - $\cos\theta$: -1.0, 0.6, 0.7, 0.8, 0.85, 0.9, 0.92, 0.94, 0.96, 0.98, 0.99, 1.0

The data samples created from these cut criteria are shown in Figure 6.1. Table 6.2 gives the number of data events in the 0–30 GeV muon momentum region for the three samples and the CC inclusive total sample.

Table 6.2.: Number of data events in the three subsamples and the inclusive sample at ND280.

CC0π	CC1π	CCoth	CCInc
17369	4047	4173	25589

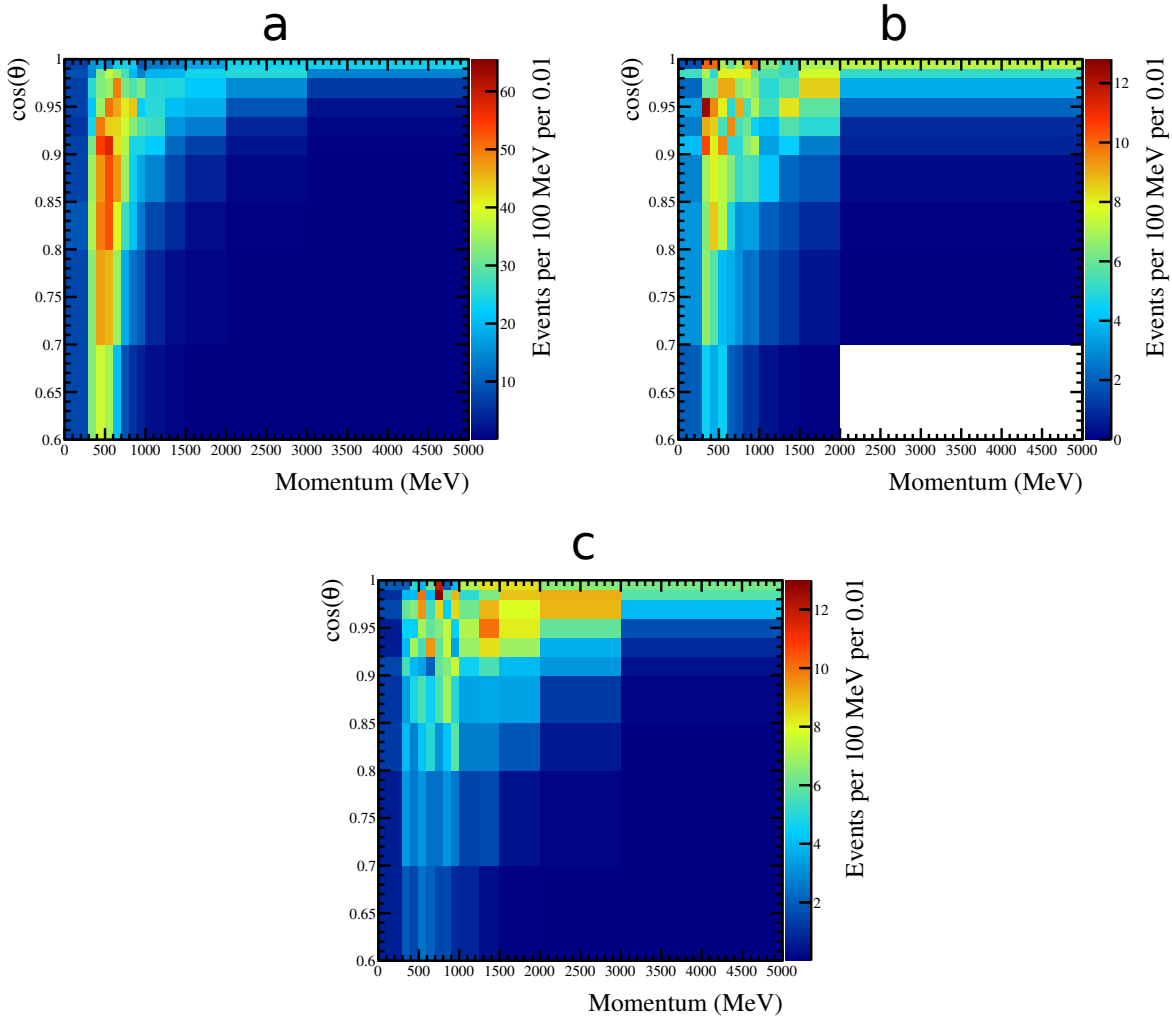


Figure 6.1.: The data samples for this analysis. Shown in (a) is the CC0π sample, in (b) the CC1π sample, and in (c) the CCoth sample.

6.3.2. SK 1R$_e$ and 1R$_\mu$

The selection for the SK data samples is described in [95]. In general, the event selection requires either a muon- or an electron-like signal, with a single ring reconstructed. The SK samples will be referred to as 1R$_\mu$ for the single ring muon-like sample (ν_μ disappearance) and likewise 1R$_e$ for the single ring electron-like sample (ν_e appearance).

For the 1R$_e$ events, the selection is as follows:

1. Fully-contained fiducial volume
2. One ring found by the ring counting algorithm
3. The ring is identified as electron-like by the PID algorithm
4. Visible energy (E_{vis}) is greater than 100 MeV
5. Zero decay electrons
6. Reconstructed neutrino energy (E_{rec}) is less than 1250 MeV
7. fitQun π^0 cut of $\ln(L_{\pi^0}/L_e) < 175 - 0.875 \times m_{\pi^0}$ (see Section 6.3.2)

There are 28 total events in this sample, selected from the Run 1-4 data, which are binned in 50 MeV-width bins from 0–1.25 GeV.

For the 1R$_\mu$ events, the selection is as follows:

1. Fully-contained fiducial volume event
2. One ring found by the ring counting algorithm
3. The ring is identified by the PID as muon-like
4. Reconstructed momentum is greater than 200 MeV/c
5. Number of decay electrons is equal to or less than one

There are 120 total events in this sample, selected from the Run 1-4 data, which are binned in 50 MeV-with bins from 0–3 GeV, then in bins with edges at 3.25, 3.5, 3.75, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0, 9.0, and 30.0 GeV.

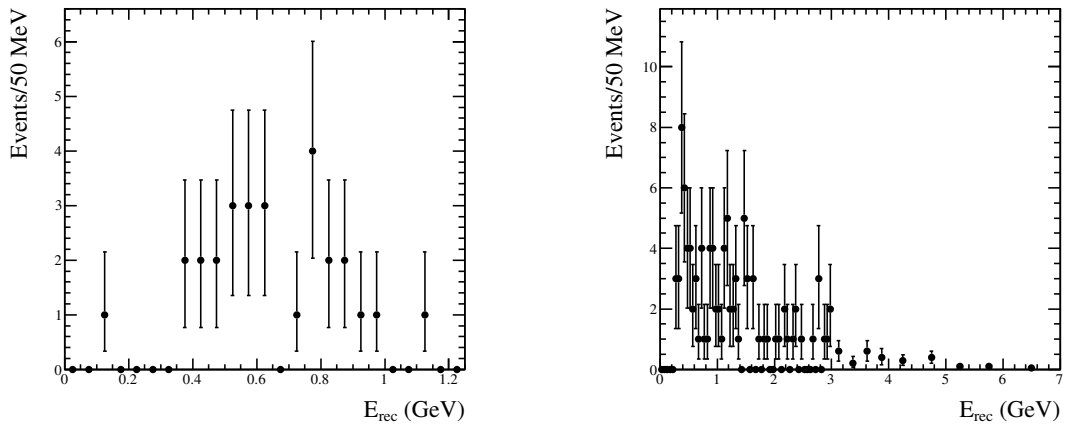


Figure 6.2.: SK data samples for Runs 1–4. Left plot shows $1R_e$ and right plot $1R_\mu$. The fit window for the $1R_\mu$ events extends to 30 GeV, but no events are found above 7 GeV, so the data is only shown up to this limit for clarity.

π^0 Rejection Cut

The biggest background for the $1R_e$ sample is $\pi^0 \rightarrow \gamma + \gamma$, where an undetectable π^0 decays into 2 photons. If the opening angle between the two photons is small, or one photon fails to be reconstructed (for example, one photon is low energy), the event will be misidentified as an electron. An event in SK is recorded via charge and time information for each PMT. This information can be used to construct a PDF for the charge and time components of an event.

A single track in the detector can be parameterized by a particle type and 7 kinematic variables; vertex position (x, y, z, t), track momentum (p) and track direction (θ, ϕ). This information can be used to construct a joint likelihood:

$$L(x) = \prod_{unhit} P(i_{unhit}|\vec{x}) \prod_{hit} P(i_{hit}|\vec{x}) f_q(q_i|\vec{x}) f_t(t_i|\vec{x}) \tag{6.2}$$

where the particle type and kinematic variables are the vector $\vec{x}$, $f_q(q_i|\vec{x})$ is the charge PDF, $f_t(t_i|\vec{x})$ is the time PDF, and i_{hit}/i_{unhit} are the PMT indices for PMTs containing hits / no hits.

The likelihood is then maximised to determine $\vec{x}$, thus reconstructing the event. This reconstruction method has been named *fiTQun* [96].

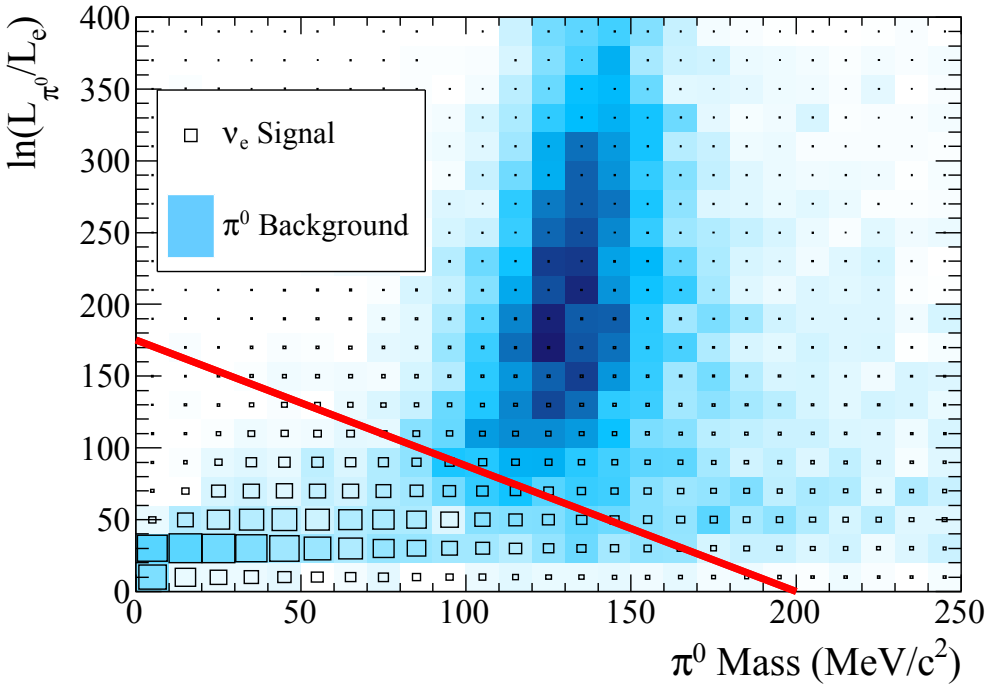


Figure 6.3.: The effectiveness of the π^0 rejection cut illustrated on signal and background events generated with MC. The red line indicates the cut threshold. The z-axis is number of MC events. Figure taken from [8].

In the context of creating a π^0 background cut, both the photon decay rings are reconstructed using this method, considering the conversion length of each photon. This method is able to locate low energy photon rings. The π^0 invariant mass can be reconstructed using the kinematics of the photons. Combining the π^0 invariant mass with the likelihood ratio between the π^0 and electron hypothesis, a π^0 cut can be determined, as shown in Figure 6.3. The resulting rejection cut reduces the SK $1R_e$ background to $\frac{1}{3}$ compared to the standard selection method, and was successfully used in T2K's appearance measurement [8].

6.4. Systematic Uncertainties

Described next are the systematic uncertainties which are to be marginalized by the MCMC. Although the analysis constructs PDFs via the event-by-event method, most systematic uncertainties are provided in a binned configuration. A discussion of the prior distributions chosen for each parameter is also given.

6.4.1. Flux

The flux prediction was described in Section 2.6. The neutrino beam flux systematic uncertainty is determined from hadron production and beamline component uncertainty, as described in [97] and [59]. The flux uncertainty is provided as a covariance matrix and implemented as normalization terms in ranges of true neutrino energy and true neutrino flavour. This means that these parameters scale the number of events for a specific range of true neutrino energy and flavour for each MC prediction. This is performed in an event-by-event manner.

The fractional covariance matrix is binned in 11 bins for ν_μ, 5 bins for $\bar{\nu}_\mu$, 7 bins for ν_e, and 2 bins for $\bar{\nu}_e$ for both ND280 and SK as follows, in true neutrino energy (GeV):

- ν_μ: 0.0, 0.4, 0.5, 0.6, 0.7, 1.0, 1.5, 2.5, 3.5, 5.0, 7.0, 30.0
- $\bar{\nu}_\mu$: 0.0, 0.7, 1.0, 1.5, 2.5, 30.0
- ν_e: 0.0, 0.5, 0.7, 0.8, 1.5, 2.5, 4.0, 30.0
- $\bar{\nu}_e$: 0.0, 2.5, 30.0

The binning scheme was chosen to have finer binning around the oscillation maximum ($\sim$0.6 GeV). Figure 6.4 shows the size of the assumed errors in the fractional covariance matrix. All flux parameters have a Gaussian prior with width given by the standard deviation extracted from the covariance matrix, and a mean value of 1.

6.4.2. Cross Section

The modelling of neutrino interactions in T2K is done with NEUT, as described in Section 2.6.3. Analyses of external data by the *neutrino interactions working group* (NIWG) are used to tune some cross section parameters used by NEUT, and determine their prior uncertainties [8]. The parameterizations of the cross section uncertainties are given in Table 6.3. The priors for all parameters are independent from one another, except for M_A^{RES}, CC1π E1, and NC1π^0, which have correlations between them as discussed in [98], which is where the cross section model used for T2K analyses is detailed. The cross section parameters are either normalization parameters, which affect the number of events in the MC prediction linearly, or described by response functions (detailed in Section 4.1.4) which have a non-linear effect on the number of predicted events. Some parameters apply only to either ND280 or SK, whereas some are correlated between both

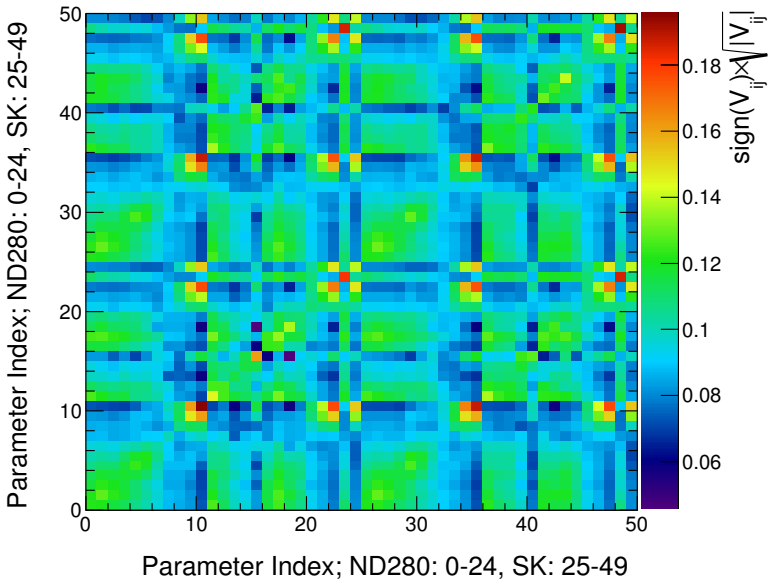


Figure 6.4.: The flux fractional covariance matrix used in the analysis. The bin indices are as follows: ND280 ν_μ (0-10), ND280 $\bar{\nu}_\mu$ (11-15), ND280 ν_e (16-22), ND280 $\bar{\nu}_e$ (23-24), SK ν_μ (25-35), SK $\bar{\nu}_\mu$ (36-40), SK ν_e (41-47), and SK $\bar{\nu}_e$ (48-49), with the energy divisions for the neutrino types given in the text.

detectors. Parameters labelled ^{12}C refer to the carbon target medium in ND280, and ^{16}O refers to the oxygen target in the water of SK.

The M_A^{QE} and M_A^{RES} parameters control the axial mass used in the model of CCQE and CC resonant pion interaction modes respectively. Normalization parameters for CCQE, CC1$\pi^\pm$ and NC1$\pi^\pm$ interaction modes are introduced to give additional degrees of freedom for specific energy ranges.

The relativistic Fermi gas model is implemented with two parameters which are tuned to external quasi-elastic scattering data. The p_F parameter accounts for uncertainty in Fermi momentum levels, and E_b accounts for uncertainty in the nucleon binding energy.

The relativistic Fermi gas model can be replaced with an alternate spectral function (SF) model. The SF parameters control a continuous combination between both models.

The pionless delta decay parameter controls the number of Δ resonance events that interact intra-nucleus, producing an event topology with no pions that can be misreconstructed as a CCQE interaction.

The W-shape parameter controls the number of CC1$\pi^\pm$ events where the pion is reabsorbed in a final state interaction before it can be detected. The model pion momentum distribution in resonance interactions shows a disagreement with data [99], and so the W-shape parameter controls the initial pion momentum distribution in NEUT, which affects how many pions are reabsorbed in the final state.

The CC Other shape parameter controls the change in shape of MC prediction by varying the number of CC multi-pion, CC DIS, CC η, K and photon production event topologies.

Uncertainty in the cross sections of CC Coherent, NC 1π^0, NC 1$\pi^\pm$, NC Coherent and NC Other is implemented for SK with normalization parameters.

The ν_μ/ν_e and $\nu/\bar{\nu}$ parameters are normalizations that vary the cross section ratios between the stated neutrino flavours. The two types of systematic, shape (response function) and normalization, are treated differently. For the shape parameters, the treatment is different between ND280 and SK. At ND280 a spline is created for each MC event by evaluating the changed weight for a particular MC event at evenly spaced points (between 2 and 7 points, depending on the parameter) and creating a cubic spline to interpolate between the points. This spline is then evaluated for the desired reweighting

Table 6.3.: Cross section parameters for the analysis, showing the applicable range of neutrino energy, nominal value and prior uncertainty. The type of systematic (shape or normalization) is also shown. Information taken from [98].

Parameter	E_ν Range	Nominal	Error	Class
M_A^{QE}	all	1.21 GeV/c^2	0.45	shape
M_A^{RES}	all	1.41 GeV/c^2	0.11	shape
p_F ^{12}C	all	217 MeV/c	30	shape
E_B ^{12}C	all	25 MeV	9	shape
SF ^{12}C	all	0 (off)	1 (on)	shape
CC Oth shape ND280	all	0.0	0.40	shape
p_F ^{16}O	all	225 MeV/c	30	shape
E_B ^{16}O	all	27 MeV	9	shape
SF ^{16}O	all	0 (off)	1 (on)	shape
CC Oth shape SK	all	0.0	0.40	shape
W-Shape	all	0.0	0.20	shape
Pionless Delta Decay	all	0.0	0.2	shape
CCQE E1	$0 < E_\nu < 1.5$	1.0	0.11	norm
CCQE E2	$1.5 < E_\nu < 3.5$	1.0	0.30	norm
CCQE E3	$E_\nu > 3.5$	1.0	0.30	norm
CC1π E1	$0 < E_\nu < 2.5$	1.15	0.43	norm
CC1π E2	$E_\nu > 2.5$	1.0	0.40	norm
CC Coh	all	1.0	1.0	norm
NC1π^0	all	0.96	0.43	norm
NC 1$\pi^\pm$	all	1.0	0.3	norm
NC Coh	all	1.0	0.3	norm
NC other	all	1.0	0.30	norm
ν_μ/ν_e	all	1.0	0.03	norm
$\nu/\bar{\nu}$	all	1.0	0.40	norm

value of the parameter, and that weight is applied to the event. The total number of ND280 splines is $\sim$4 million, and are evaluated on a GPU as described in Section 4.1.4.

At SK, splines are created by evaluating the weights for particular MC events at evenly spaced points, as for ND280, but binning those events in E_{rec} and E_{true}, which creates an averaged reweighing for a particular kinematic bin. Each MC event is weighted according to the evaluated spline for the kinematic bin of that event. For the normalization parameters, the event is simply weighted by the value of the parameter.

In this analysis, a uniform prior restricted to the range $[0, 1]$ is used for the spectral function parameters; as a result, they do not enter the constraint term in Equation 6.1.

6.4.3. ND280 Detector

The systematic uncertainty for the ν_μ ND280 detector selections are defined by 19 parameters which are described in [100]. In this technical note, the event selection is rerun for each iteration of the fitting algorithm to model. For this analysis, the method of reanalyzing the entire ND280 sample for each MCMC step was computationally prohibitive. Therefore, a covariance matrix approach similar to the flux uncertainty was used. The fractional covariance matrix was produced by generating 2000 throws of the inputs for the detector systematics, and the full detector systematic analysis was used for each throw. The fractional covariance for each bin of the matrix was calculated as

$$V_{ij} = \frac{1}{2000} \sum_{n=1}^{2000} \frac{\left(N_n^{\text{reweighted},i} - N^{\text{average},i}\right)\left(N_n^{\text{reweighted},j} - N^{\text{average},j}\right)}{N^{\text{average},j} N^{\text{average},i}} \quad (6.3)$$

where $N^{\text{average},i}$ is the average of the 2000 throws.

The binning for the detector systematic covariance matrix was chosen to be coarser than the binning used for fitting the data, in order to reduce the number of parameters used in the fit, especially as the size of the detector systematic errors is typically smaller than the size of the flux and cross section errors. The binning chosen for all samples has seven bins in muon momentum and five bins in $\cos\theta$ and is as follows:

- p_μ (MeV): 0, 300, 500, 600, 700, 1000, 2000, 30000

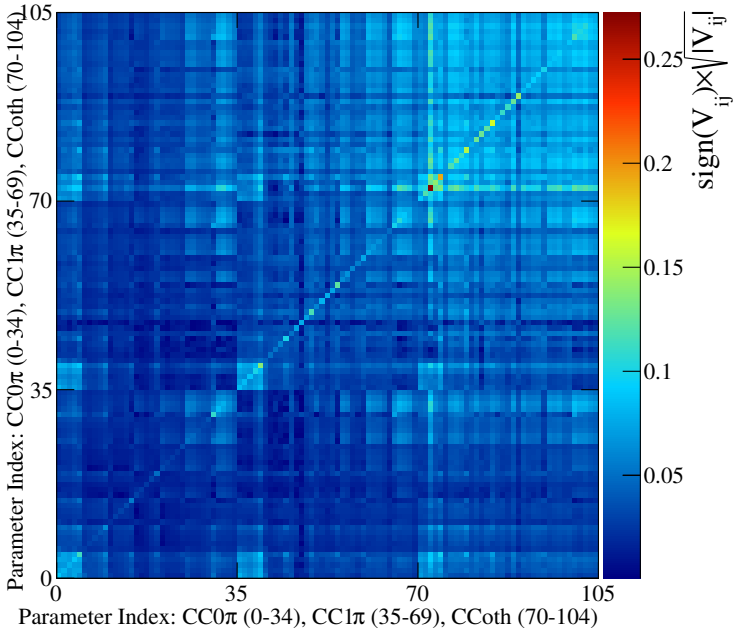


Figure 6.5.: The detector fractional covariance matrix used in the analysis. In this matrix, bins 0-34 cover the CC0π sample, 35-69 the CC1π sample, and 70-104 the CC other sample. Within each sample, the bins iterate over $\cos\theta$ from low to high for the lowest momentum bin, then from low to high for the second lowest momentum bin, etc.

- $\cos\theta$: -1, 0.85, 0.9, 0.94, 0.98, 1.0

The fractional covariance matrix is shown in Figure 6.5. In this matrix, bins 0-34 cover the CC0π sample, 35-69 the CC1π sample, and 70-104 the CCoth sample. Within each sample, the bins iterate over $\cos\theta$ from low to high for the lowest momentum bin, then from low to high for the second lowest momentum bin, etc.

To apply this systematic, each event is weighted by the value according to the bin corresponding to the event's reconstructed momentum and angle.

6.4.4. Final State Interactions (ND280 only)

Final state interaction (FSI) uncertainties manifest from true event topologies being masked by interactions that take place by the final state particles before they are detected. For this analysis, the six FSI parameters used are described in [98], which are scaling parameters that affect pion production, pion absorption, low and high energy charge exchange and low and high energy inelastic interactions in NEUT. The covariance of the parameters are determined as follows.

The parameters are varied simultaneously and a 1 σ surface is constructed by comparing to external π^+-^{12}C scattering data (see Appendix in [101]). This surface is then sampled by generating 16 parameter sets seen in Table 6.4, which were placed such that there is a parameter configuration spanning each octant of two 3-dimensional spaces; low and high energy parameters respectively.

Par. Set	FSIQE	FSIQEH	FSIINEL	FSIABS	FSICX	FSICXH
Nominal	1.0	1.8	1	1.1	1.0	1.8
15	0.6	1.1	1.5	0.7	0.5	2.3
16	0.6	1.1	1.5	0.7	1.6	2.3
17	0.7	1.1	1.5	1.6	0.4	2.3
18	0.7	1.1	1.5	1.6	1.6	2.3
19	1.4	1.1	1.5	0.6	0.6	2.3
20	1.3	1.1	1.5	0.7	1.6	2.3
21	1.5	1.1	1.5	1.5	0.4	2.3
22	1.6	1.1	1.5	1.6	1.6	2.3
23	0.6	2.3	0.5	0.7	0.5	1.3
24	0.6	2.3	0.5	0.7	1.6	1.3
25	0.7	2.3	0.5	1.6	0.4	1.3
26	0.7	2.3	0.5	1.6	1.6	1.3
27	1.4	2.3	0.5	0.6	0.6	1.3
28	1.3	2.3	0.5	0.7	1.6	1.3
29	1.5	2.3	0.5	1.5	0.4	1.3
30	1.6	2.3	0.5	1.6	1.6	1.3

Table 6.4.: Parameter sets generated from the 1 σ surface created by comparing to MiniBooNE data. The 16 variations were chosen to span the extent of the parameter space, so that the covariance can be determined. Information taken from [98].

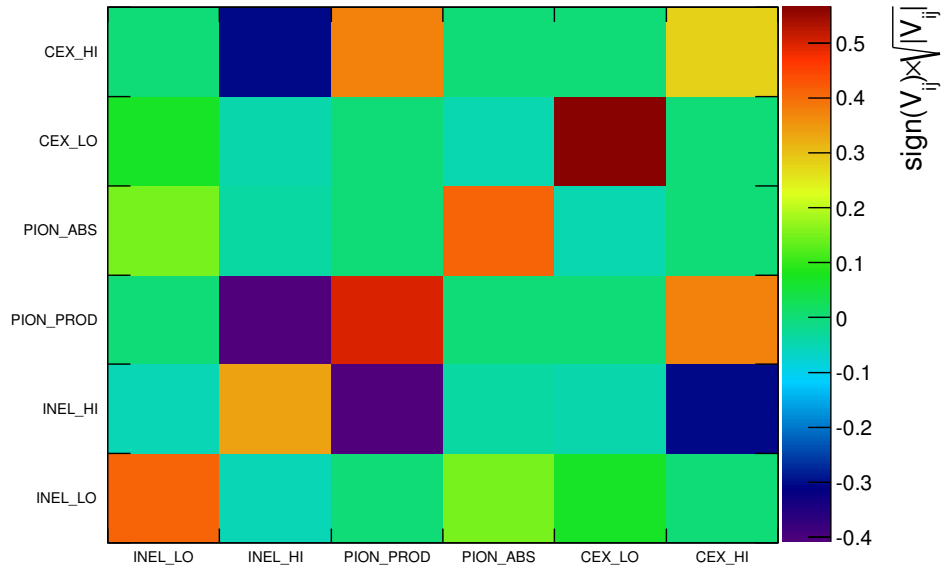


Figure 6.6.: The FSI covariance matrix used in the analysis. The parameters are defined in [98].

The covariance between the parameter sets in Table 6.4 was used to construct a covariance matrix, shown in Figure 6.6. This covariance matrix is used as a multivariate prior for the ND280 FSI parameters.

The response of the PDF when tweaking each parameter was modelled by creating a spline for each MC event. This spline is then evaluated for the desired reweighting value of the parameter, and that weight is applied to the event when constructing the PDF.

6.4.5. SK Detector

The SK detector systematics are correlated between the $1R_e$ and $1R_\mu$ samples, as described in [102]. The SK detector systematic uncertainties are determined by an analysis of atmospheric neutrino control samples from SK, along with a hybrid-π^0 sample and cosmic ray data, which are combined into a unified selection for ν_μ and ν_e.

The uncertainties on topological selection criteria as described in Section 6.3 are determined using the control samples and by shifting the selection parameters simultane-

ously in a MCMC framework. The uncertainties and posterior parameter covariance are determined from the marginalized posterior distribution.

The uncertainty on the selection parameters are propagated to the oscillation analyses using a toy MC to convert the error into an uncertainty on ranges of kinematic variables such as E_{rec} and p-θ. For this analysis, the PDFs are created in E_{rec} space, so the error matrix for E_{rec} is used.

The first 12 parameters are for the $1R_e$ sample, in four sets of three energy bins (0–0.35; 0.35–0.8; 0.8–1.25 GeV) for the signal ν_e, beam ν_μ CC, beam ν_e CC, and NC event topologies. The following 6 parameters are for $1R_\mu$: three energy bins (0–0.4; 0.4-1.1; 1.1-30 GeV) for ν_μ CCQE, one bin for ν_μ CCnQE, one bin for ν_e CC, and one bin for NC events. The final bin is the energy scale uncertainty.

The energy scale uncertainty receives a different treatment from the normalization parameters. Instead of a weight, it is applied as a multiplicative factor for E_{rec}, effectively scaling the MC prediction to higher and lower energies. This method of applying a direct scaling factor to the kinematic variable of the PDF is a benefit of the event-by-event reweighting method, and is much more difficult to implement if trying to apply the energy scale to histogram, as is done in other T2K analyses.

The fractional covariance matrix is shown in Figure 6.7. The matrix contains the final state interaction + secondary interaction (FSI+SI) uncertainties for SK, which are calculated in a similar way to the ND280 FSI (Section 6.4.4). As these matrices are provided in the same binning format as for the detector systematics, they are summed linearly into a single matrix.

6.5. Monte Carlo Predictions and Pre-fit Data/MC comparison

The MC for both detectors is reweighted to predict event rates for nominal values of systematic parameters. In the case of SK, the MC is also weighted according to the oscillation probability.

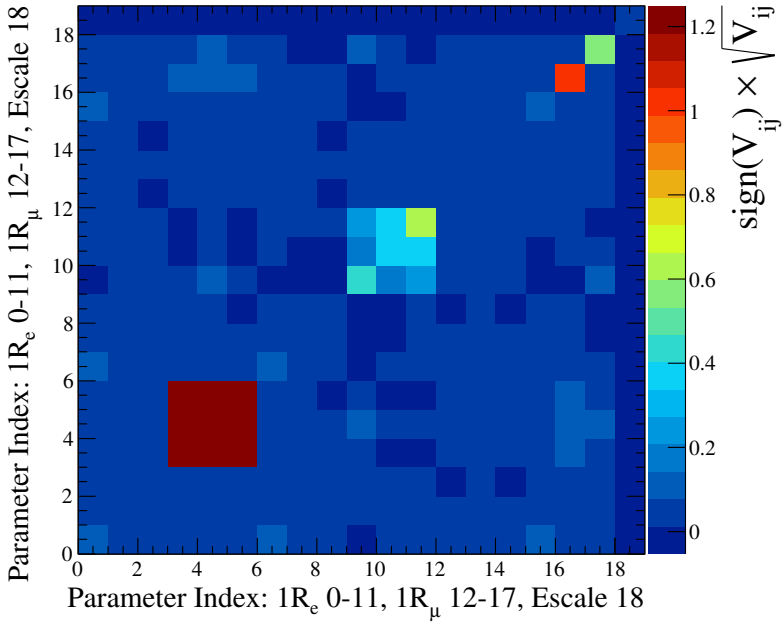


Figure 6.7.: The SK detector fractional covariance matrix used in the analysis. The errors for $1R_e$ are in bins 0-11, $1R_\mu$ in bins 12-17, and the energy scale error in bin 18.

6.5.1. Beam and Near Detector Extrapolation

A separate analysis of ND280 data exists with the intention of constraining correlated flux and cross section parameters by producing a covariance matrix and extrapolating the result as an input to oscillation analyses. This analysis is named the *Beam and Near detector Flux Fit* (BANFF), as described in [8] and [32]. In the analysis presented in this thesis, the extrapolation method is replaced with a simultaneous analysis of both SK and ND280 data. The BANFF results can be used as a cross-check of this analysis, and vice-versa. This analysis uses the parameterization described by the neutrino interactions working group [98] (NIWG), which means that analyses using the nominal values produced by the BANFF fit will have different MC predictions.

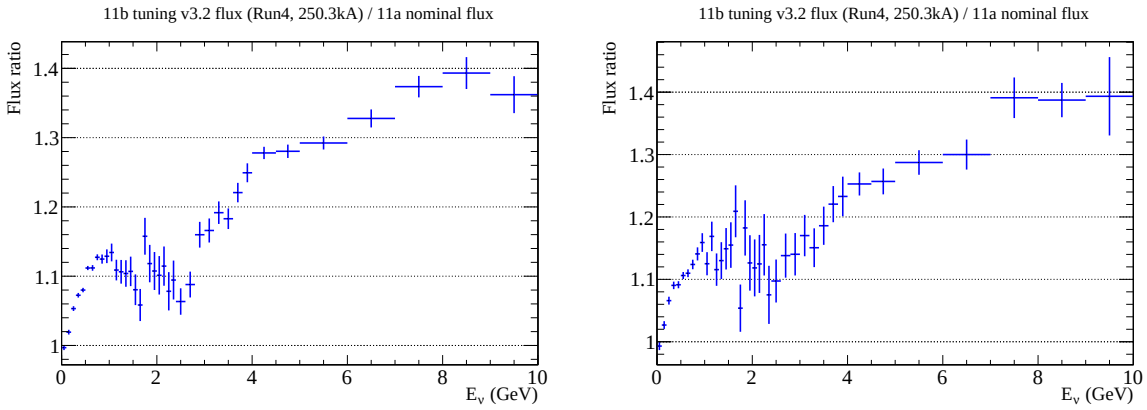
For clarification, when parameter nominal values are set to those recommended by the NIWG, they will be referred to as the NIWG tuning. When the parameter nominal values are set to those used by the BANFF fit, they will be referred to as the BANFF tuning. The analysis uses the NIWG tuning, but some event rates and plots will be presented using the BANFF tuning for validation purposes only.

6.5.2. Flux Tuning

The MC for both detectors is generated using the beam flux simulation as described in Section 2.6. However, updates to the beam simulation require that the MC also be updated. Due to the computational expense of regenerating the detector MC, instead flux reweighting histograms are provided. These reweighting factors change the shape and normalization of the MC predictions to better reflect the beam flux simulation without having to regenerate the MC. An example of the flux correction histograms can been seen in Figure 6.8.

6.5.3. ND280

The nominal ND280 MC undergoes two tunings to generate the initial predicted distributions. First, the events are weighted according to the updated flux simulation tuning as described in Section 6.5.2. Second, the events are weighted for the values of the cross section parameters M_A^{RES}, CC1π E1, and NC1π0 according to a fit to the MiniBooNE CC1π data as described by the NIWG in [98], as these values are different from the



(a) SK **(b)** ND280

Figure 6.8.: Flux tuning for 6.8a SK and 6.8b ND280 ν_μ MC samples. The histograms contain the ratio between the new and old tuning.

Table 6.5.: Number of data events in the three subsamples and the inclusive sample at ND280.

	CC0π	CC1π	CCoth	CCInc
Data	17369	4047	4173	25589
Pre-fit MC	19980	5037.2	4729.1	29746.1
Data/MC Ratio	0.869	0.803	0.882	0.860

values of these parameters used to generate the MC. Table 6.5 gives the number of events in the 0–30 GeV/c muon momentum region for the data and the MC. After these two reweightings, the MC prediction is described as the "prefit" MC prediction.

The prefit MC prediction for the data from ND280 is shown in Figure 6.9. Projections of the data and prefit MC in momentum and angle are shown in Figure 6.10. Generally, the MC predicts a larger number of events than the data, with the effect more pronounced in the CC0π and CC1π samples than in the CCoth sample.

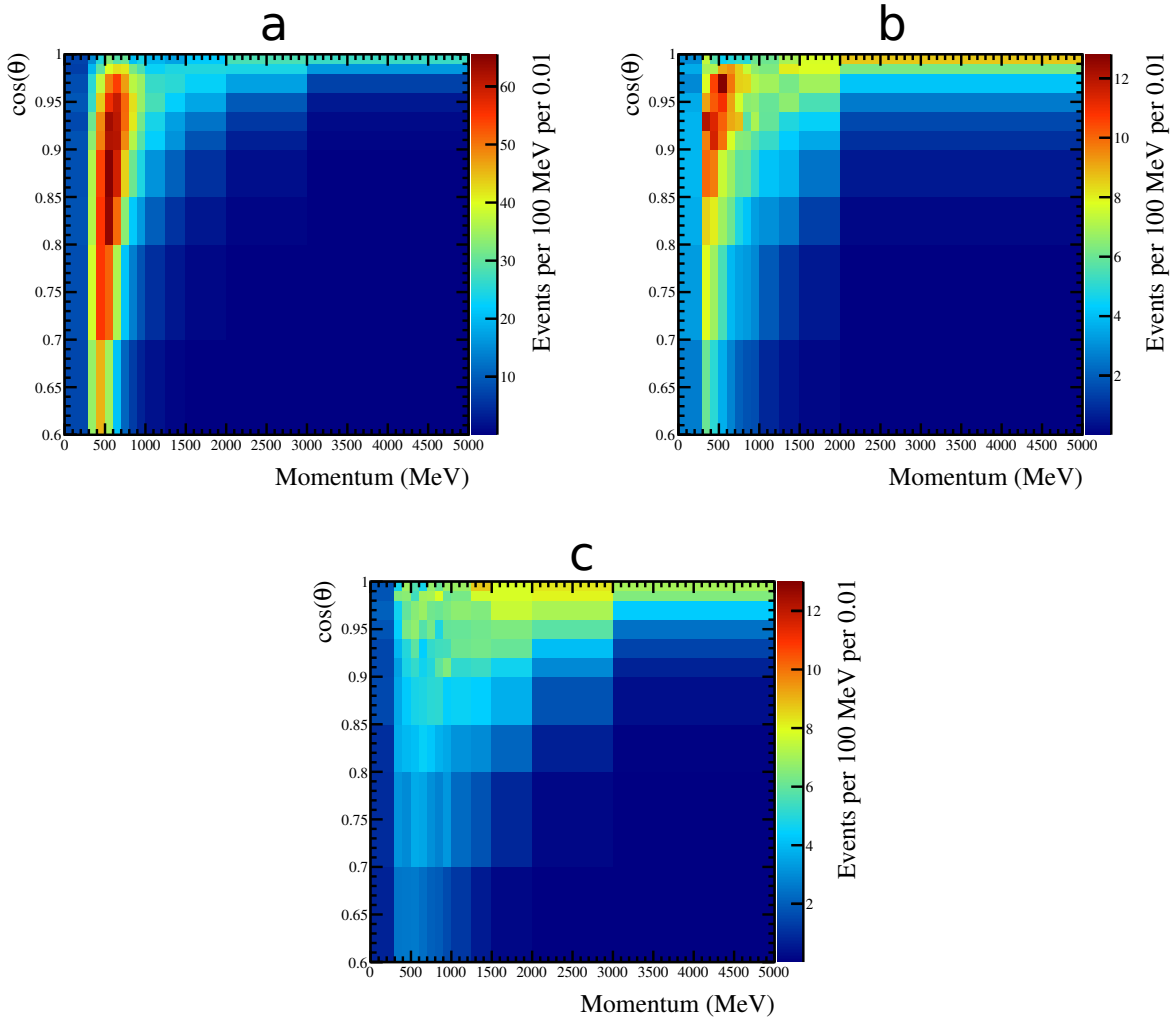


Figure 6.9.: The nominal number of MC predicted events in the p–cos θ binning used for the fit. The highest momentum and backwards angle bins are not shown for clarity. Shown in (a) is the CC0π sample, in (b) the CC1π sample, and in (c) the CCoth sample.

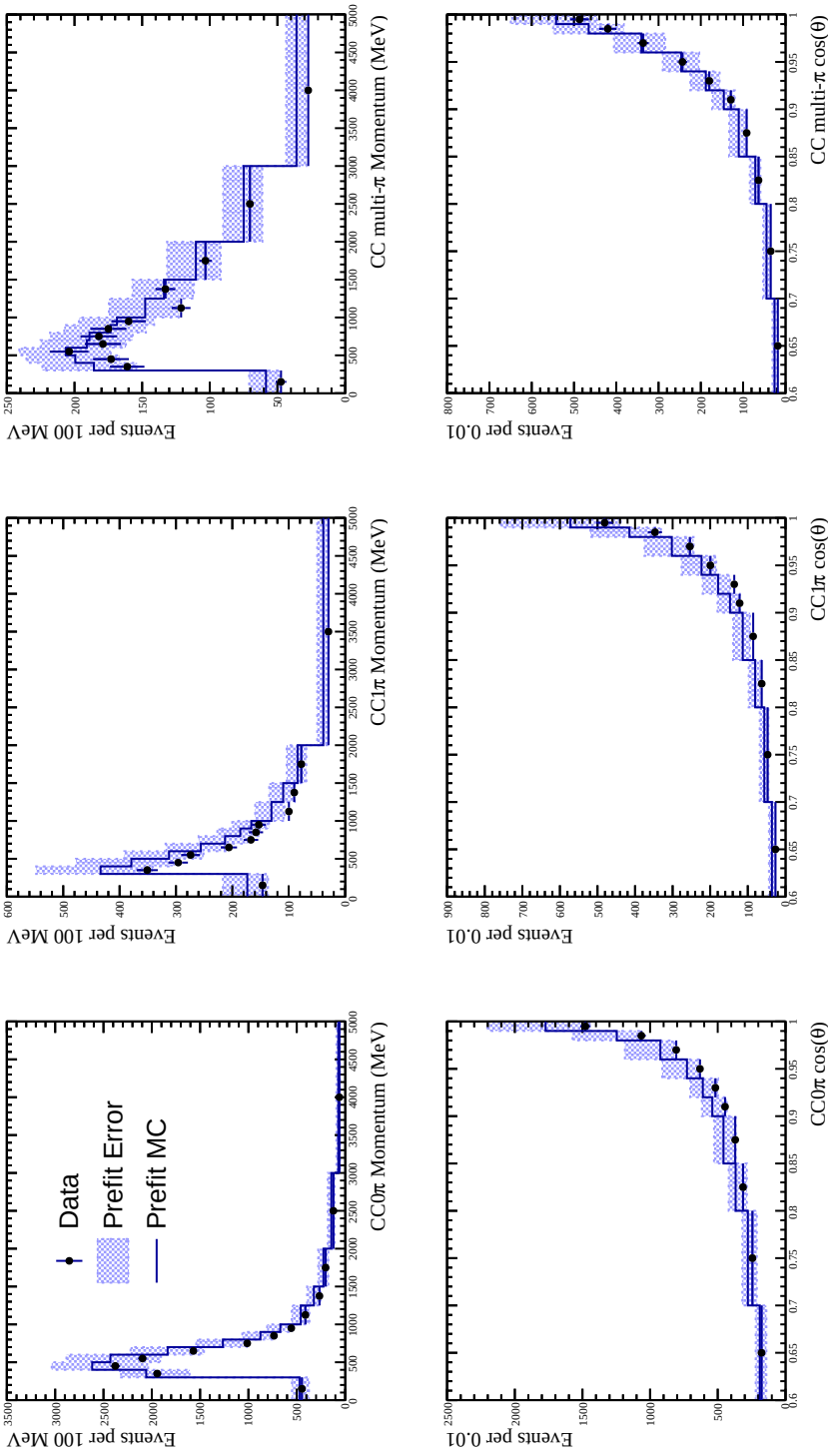


Figure 6.10.: The data and predicted number of MC events projected onto the momentum (top row) and $\cos\theta$ axes. The left, middle and right columns correspond to the CC0π, CC1π and CC other samples. Error bands are produced by drawing many random throws from the prior distribution for each systematic parameter, and observing the variance in each kinematic bin of momentum and angle.

6.5.4. Super-Kamiokande

SuperK detector MC is provided by the T2K-SK working group, which is used to create the PDFs used in the far detector fit. The MC is provided in 5 categories:

- ν_μ signal
- ν_e signal
- ν_e background
- $\vec{\nu}_\mu$ background
- $\vec{\nu}_e$ background

where ν_e signal is created by combining the ν_μ flux with the ν_e interaction cross-section, to create a sample that represents maximal oscillation of $\nu_\mu \rightarrow \nu_e$. This sample is reduced with oscillation reweighting. After passing the $1R_\mu$ and $1R_e$ selection cuts as prescribed in Section 6.3.2, the samples are normalized to 6.57×10^{20} protons on target, and then have the following oscillation probability weights applied:

- $P(\nu_\mu \rightarrow \nu_\mu)$ for ν_μ signal
- $P(\nu_\mu \rightarrow \nu_e)$ for ν_e signal
- $P(\nu_e \rightarrow \nu_e)$ for ν_e background
- $P(\vec{\nu}_\mu \rightarrow \vec{\nu}_\mu)$ for $\vec{\nu}_\mu$ background, and
- $P(\vec{\nu}_e \rightarrow \vec{\nu}_e)$ for $\vec{\nu}_e$ background.

The samples are separated logically into the following 9 neutrino interaction categories to facilitate the application of systematic uncertainties:

- Charged Current Quasi Elastic (CCQE)
- Charged Curent Single Pion Resonant Production (CC$1\pi^\pm$)
- Charged Current Coherent Pion Production (CC Coh)
- Charged Current Multi-Pion Resonant Production (CC nπ)
- Charged Current Other (CC Other)
- Neutral Current π^0 (NCπ^0)

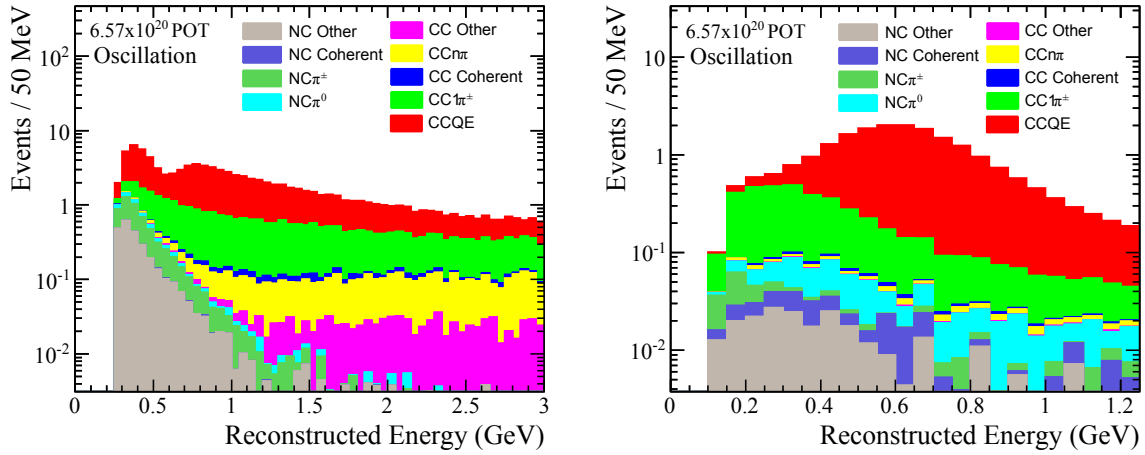


Figure 6.11.: $1R_\mu$ and $1R_e$ MC selections, separated into neutrino interaction mode type. Weighted for oscillation hypothesis: $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0251$, $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$.

- Neutral Current Single Pion Resonant Production ($NC\pi^\pm$)
- Neutral Current Coherent Pion Production (NC Coh)
- Neutral Current Other (NC other).

These modes are illustrated in Figures 6.11 and 6.12. The remaining treatment of the $1R_\mu$ and $1R_e$ samples are described in the following sections.

SK $1R_\mu$

The nominal MC undergoes two tunings to generate the initial predicted distributions. First, the events are weighted according to the updated flux predictions as described in Section 6.5.2. Secondly, the events are weighted for the values of the cross section parameters M_A^{RES}, CC1π E1, and NC1π0 according to a fit to the MiniBooNE CC1π data as described by NIWG in [98], and demonstrated in Figure 6.13. Table 6.6 gives the number of events in the 0–30 GeV reconstructed energy range, broken down by sample type and interaction mode. The effect of applying the flux correction and NIWG tuning is shown in Figure 6.14. Additionally, using instead the BANFF tuning as described in Section 6.5.1, the MC prediction finds a rate of 125 events for the oscillation hypothesis and 446 events for the null oscillation hypothesis (both hypotheses are described in the caption of Table 6.6).

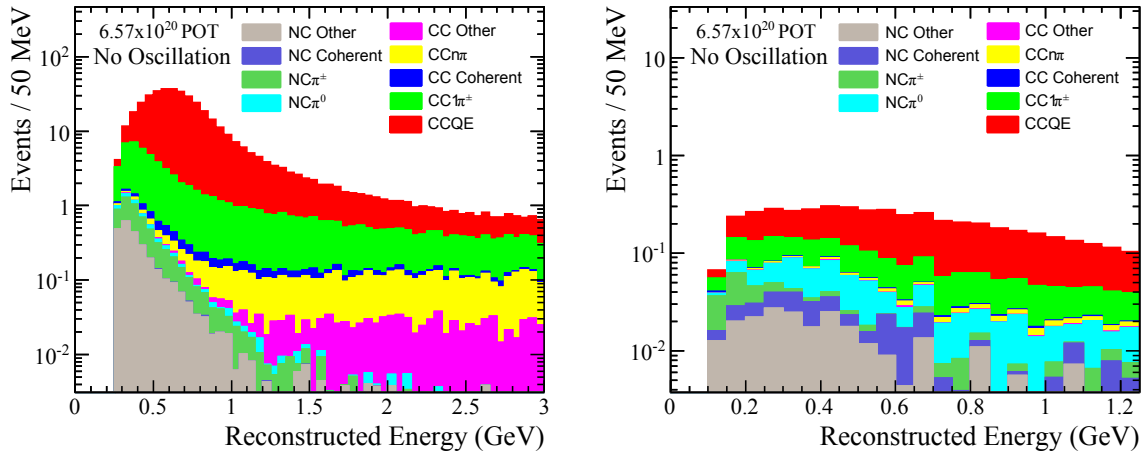
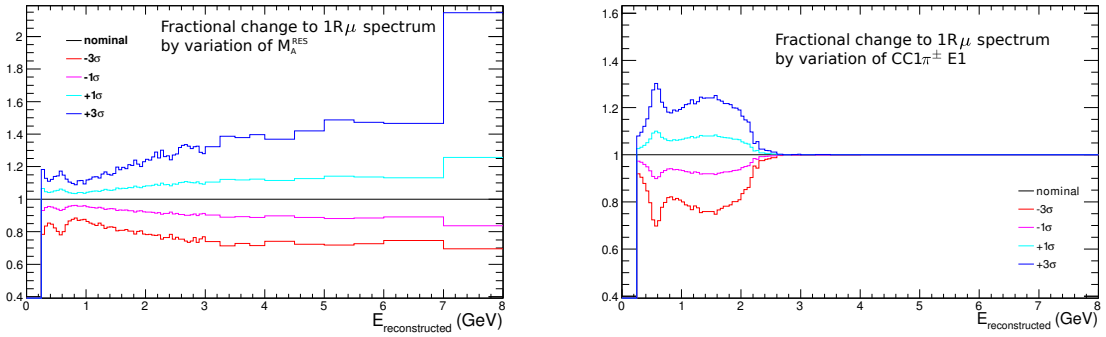


Figure 6.12.: $1R_\mu$ and $1R_e$ MC selections, separated into neutrino interaction mode type. Weighted for null oscillation hypothesis: $\sin^2\theta_{23} = 0.0$, $\sin^2\theta_{13} = 0.0$, $\sin^2\theta_{12} = 0.0$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$.



(a) Effect on the $1R_\mu$ prediction by applying $\pm$ 1 and 3σ tweaks to the M_A^{RES} parameter.

(b) Effect on the $1R_\mu$ prediction by applying $\pm$ 1 and 3σ tweaks to the CC1$\pi^\pm$ E1 parameter.

Figure 6.13.: Sigma variations to M_A^{RES} and CC1$\pi^\pm$ E1 and the fractional effect on the $1R_\mu$ prediction. The M_A^{RES} parameter has a non-linear response described by splines, whereas CC1$\pi^\pm$ E1 is a linear parameter described by a parameter in a covariance matrix.

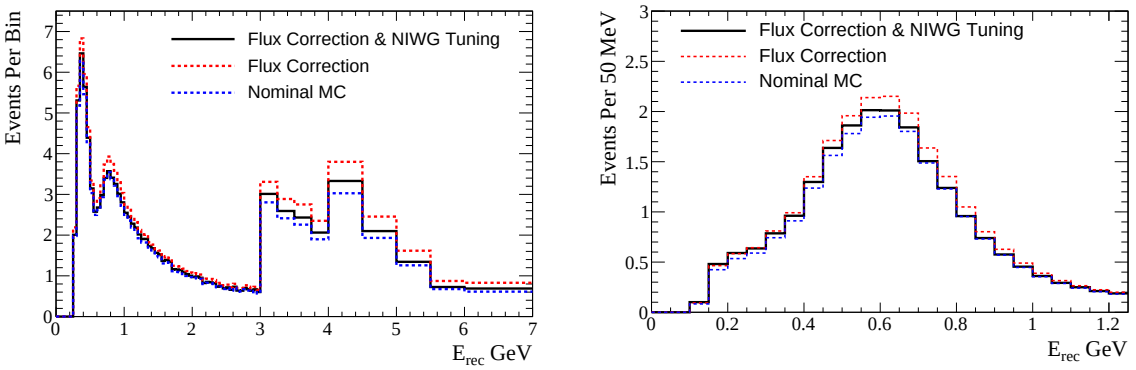
Table 6.6.: Top: Oscillated rates for $1R_\mu$, tuned by NIWG for 6.57×10^{20} POT. Oscillation parameters used: $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0251$, $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$. Bottom: Unoscillated rates for $1R_\mu$. All mixing angles set to zero.

	ν_μ	ν_e	$\bar{\nu_\mu}$	$\bar{\nu_e}$	$\nu_e$signal
CCQE	73.583	0.035	4.782	0.002	0.198
CC1π	41.398	0.029	2.949	0.002	0.081
CC coherent	0.897	0.001	0.247	0.000	0.005
CCnπ	6.558	0.004	0.404	0.000	0.001
CC other	2.175	0.003	0.100	0.000	0.001
NCπ^0	0.945	0.032	0.054	0.004	0.000
NC$\pi^{+/-}$	4.638	0.131	0.262	0.016	0.000
NC coherent	0.018	0.000	0.001	0.000	0.000
NC other	2.764	0.112	0.158	0.012	0.000
Sample Totals	132.977	0.348	8.956	0.036	0.285
Total Rate	142.603				

	ν_μ	ν_e	$\bar{\nu_\mu}$	$\bar{\nu_e}$	$\nu_e$signal
CCQE	367.066	0.038	9.710	0.002	0.000
CC1π	81.343	0.031	4.143	0.002	0.000
CC coherent	2.138	0.001	0.462	0.000	0.000
CCnπ	7.465	0.004	0.461	0.000	0.000
CC other	2.304	0.003	0.107	0.000	0.000
NCπ^0	0.945	0.032	0.054	0.004	0.000
NC$\pi^{+/-}$	4.638	0.131	0.262	0.016	0.000
NC coherent	0.018	0.000	0.001	0.000	0.000
NC other	2.764	0.112	0.158	0.012	0.000
Sample Totals	468.681	0.353	15.358	0.036	0.000
Total Rate	484.428				

SK $1R_e$

The nominal MC undergoes two tunings to generate the initial predicted distributions. First, the MC events are weighted according to the updated flux predictions as described in Section 6.5.2. Second, the MC events are weighted for the values of the cross section parameters M_A^{RES}, CC1π E1, and NC1π0 according to a fit to the MiniBooNE CC1π



(a) $1R_\mu$ **(b)** $1R_e$

Figure 6.14.: Effect of applying the flux correction and NIWG tuning to the $1R_\mu$ and $1R_e$ MC predictions.

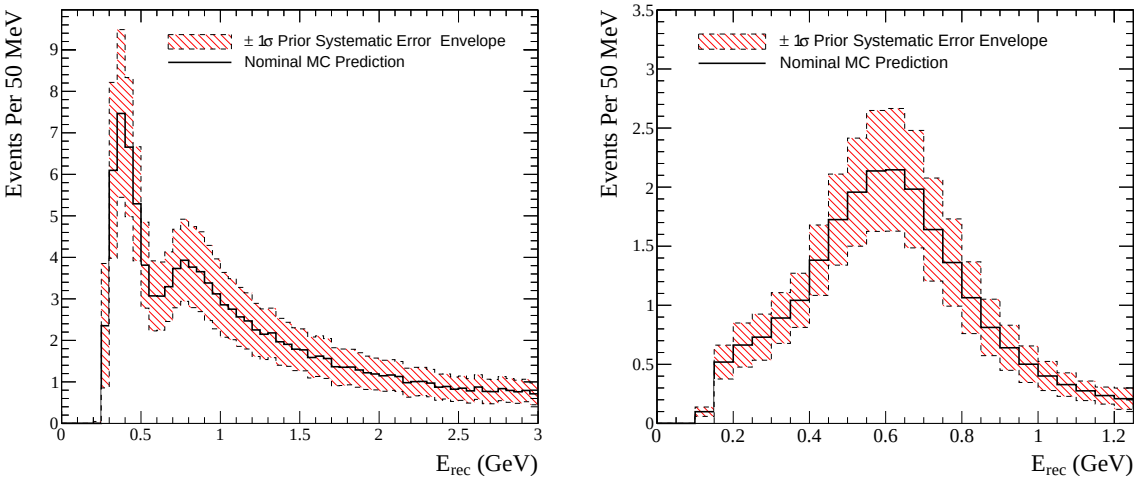
data as described by NIWG in [98], and demonstrated in Figure 6.13. Table 6.7 gives the number of events in the 0–1250 MeV reconstructed energy range, broken down by sample type and interaction mode. The effect of applying the flux correction and NIWG tuning is shown in Figure 6.14. Additionally, using instead the BANFF tuning as described in Section 6.5.1, the MC prediction finds a rate of 21.6 events for the oscillation hypothesis (described in the caption of Table 6.7) and 4.9 events for the null oscillation hypothesis ($\sin^2\theta_{13} = 0.0$).

Table 6.8 shows the number of data events and predicted prefit MC events and their ratios for the two samples, using PDG2012 values [103] for the oscillation parameters; Figure 6.16 shows the same graphically as a function of E_{rec}, along with the unoscillated spectra. Figure 6.17 shows a scan of the total predicted rates as a function of oscillation parameters. Figure 6.15 shows the total prior systematic uncertainty on the $1R_\mu$ and $1R_e$ MC predictions. The envelopes were generated by throwing 10000 sets of systematic parameters from the prior distributions. A histogram is created for each bin in the spectra. For each throw, the MC prediction is generated, and the event rate for each bin is added to the corresponding histogram. Then, the per-bin histograms are fit with a Gaussian distribution, whose mean value is taken to be the nominal event rate for that bin, and the sigma is the uncertainty.

Table 6.7.: Top: Oscillated rates for $1R_e$, tuned by NIWG for 6.57×10^{20} POT. Oscillation parameters used: $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0251$, $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$. Bottom: Unoscillated rates for $1R_e$. Only $\sin^2 2\theta_{13} = 0.0$; other oscillation parameters remain the same.

	ν_μ	ν_e	$\bar{\nu_\mu}$	$\bar{\nu_e}$	$\nu_e$signal
CCQE	0.050	2.276	0.001	0.098	14.989
CC1π	0.021	0.952	0.000	0.053	2.970
CC coherent	0.000	0.009	0.000	0.007	0.044
CCnπ	0.001	0.050	0.000	0.003	0.030
CC other	0.000	0.008	0.000	0.000	0.002
NCπ^0	0.475	0.015	0.024	0.002	0.000
NC$\pi^{+/-}$	0.149	0.004	0.008	0.000	0.000
NC coherent	0.181	0.005	0.016	0.001	0.000
NC other	0.329	0.010	0.013	0.001	0.000
Sample Totals	1.207	3.329	0.062	0.165	18.036
Total Rate	22.798				

	ν_μ	ν_e	$\bar{\nu_\mu}$	$\bar{\nu_e}$	$\nu_e$signal
CCQE	0.050	2.471	0.001	0.104	0.365
CC1π	0.021	1.010	0.000	0.056	0.040
CC coherent	0.000	0.010	0.000	0.007	0.001
CCnπ	0.001	0.052	0.000	0.003	0.000
CC other	0.000	0.008	0.000	0.000	0.000
NCπ^0	0.475	0.015	0.024	0.002	0.000
NC$\pi^{+/-}$	0.149	0.004	0.008	0.000	0.000
NC coherent	0.181	0.005	0.016	0.001	0.000
NC other	0.329	0.010	0.013	0.001	0.000
Sample Totals	1.206	3.585	0.062	0.175	0.406
Total Rate	5.434				



(a) $1R_\mu$ error envelope **(b)** $1R_e$ error envelope

Figure 6.15.: Error envelopes for SK samples, showing the $\pm 1\sigma$ range of all systematic uncertainties on the reconstructed energy spectrum. The error envelope shown is before constraint with ND280.

Table 6.8.: Number of data events in the SK samples, with MC tuned by NIWG for 6.57×10^{20} POT. Oscillation parameters used: $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0251$, $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$.

	$1R_e$	$1R_\mu$
Data	28	120
MC	22.798	142.603
Data/MC Ratio	1.228	0.841

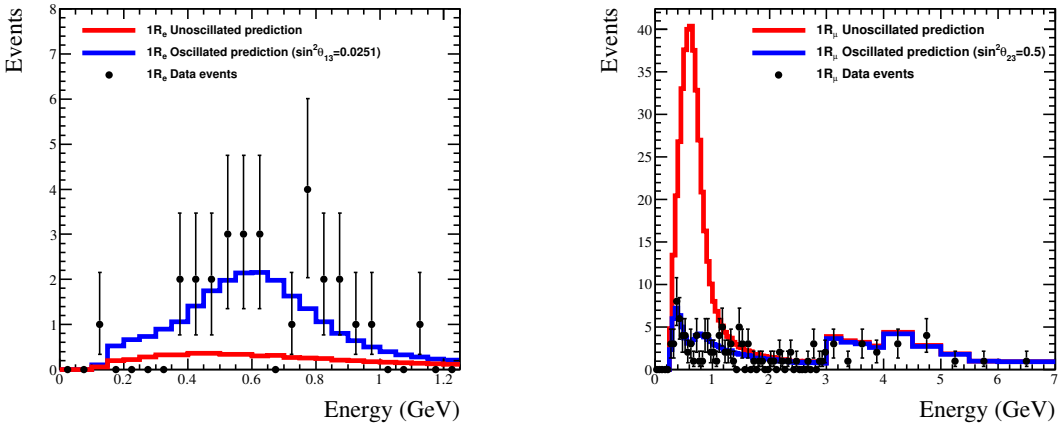


Figure 6.16.: Oscillated (blue) and unoscillated (red) spectra for $1R_\mu$ (left) and $1R_e$ (right) samples. Rates are tuned by NIWG for 6.57×10^{20} POT. Oscillation parameters used: $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0251$, $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV, $\delta_{cp} = 0$. Rates are generated with the updated flux tuning and NIWG reweighting.

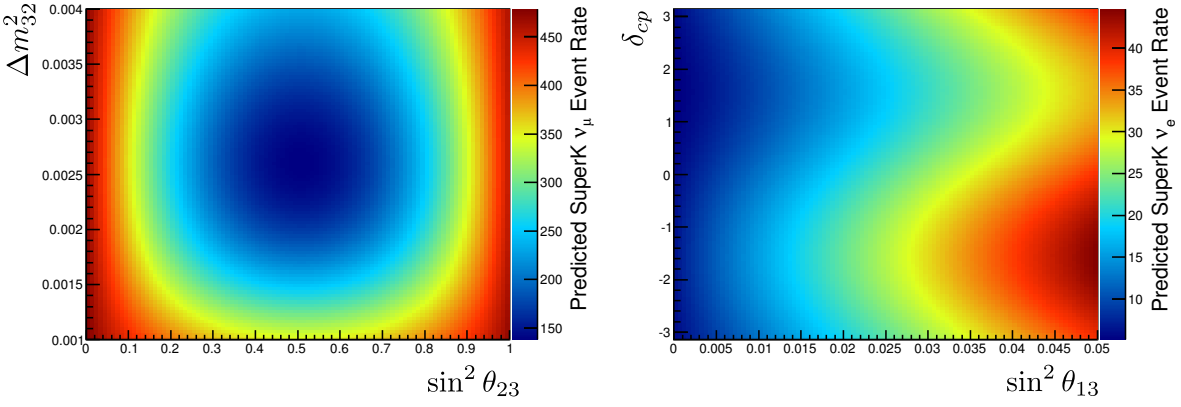


Figure 6.17.: A scan of predicted event rates for Run 1–Run 4 data of 6.57×10^{20} POT. Left shows $1R_\mu$ scanning over $\sin^2\theta_{23}$ and Δm^2_{32}; right shows $1R_e$ scanning over $\sin^2\theta_{13}$ and δ_{cp}. Other oscillation parameter are fixed at $\sin^2\theta_{23} = 0.5$ (for $1R_e$), $\sin^2\theta_{13} = 0.0251$ (for $1R_\mu$), $\sin^2\theta_{12} = 0.311$, $\Delta m^2_{12} = 7.5 \times 10^{-5}$ eV, $\Delta m^2_{32} = 2.4 \times 10^{-3}$ eV (for $1R_e$), $\delta_{cp} = 0$ (for $1R_\mu$). Rates are generated with the updated flux tuning and NIWG reweighting.

Chapter 7.

Validation and 3 Flavour Data Analysis Results

7.1. Analysis Validation

This analysis is is performed blind, meaning that analysing the Run 1–4 data is the last step to be undertaken in order to reduce bias introduced by the analyser. To prepare the analysis for unblinding, the software framework is validated via three methods: using a nominal data set in Section 7.1.1, an ensemble of toy experiments in Section 7.1.2 and a series of common fake data sets shared between T2K oscillation analysers in Section 7.1.4.

7.1.1. Nominal Data Set

A nominal dataset is defined to be a toy experiment generated from the PDF in such a way that there are no statistical or systematic fluctuation as illustrated in Figure 7.1. This is achieved by reweighting the PDF to nominal values of systematic parameters, along with the chosen oscillation parameter values, and required protons on target, but instead of drawing randomly from the PDF, the PDF is considered as the dataset. This produces a data set free from statistical fluctuations, which, when analysed, should result in parameter posterior distributions free from bias. There is still statistical uncertainty in the data set, which comes from the normalization of the PDF to POT. However, the uncertainty is centered around the true value, and so the contours produced will have a realistic width, but there will be no bias in the parameter due to statistical fluctuation. Figure 7.2 shows the results of an analysis of a nominal dataset using 20 million MCMC

steps, summarising the best fit values of all systematic parameters, their posterior error in units of σ and the true value. This plot shows minimal bias in the parameters, and is complimentary to the toy experiment results in Section 7.1.2.

Figures 7.3 and 7.4 show credible intervals and best fit values constructed from the nominal posterior distributions. Also plotted are the true parameter values of the nominal data set.

7.1.2. Toy Experiments

Toy experiments are produced by throwing mock data sets from the SK $1R_\mu$, SK $1R_e$ and the three ND280 ν_μ PDFs. This is achieved by constructing the MC prediction for nominal systematic parameter values, and using the PDG2012 values for oscillation parameters [104]. The nominal event rate is extracted by taking the integral of the MC prediction, and this is used to seed a Poisson distribution, from which a fluctuated event rate is drawn for each toy. Random draws from the PDFs are then made using the fluctuated rates.

Systematic fluctuation is introduced by throwing the central values for the systematic priors when performing the analysis, as in the method described in [105]. This method is used because if instead the systematic central value was thrown randomly when the toy data set was generated, in combination with a fluctuation of the event rate, the total uncertainty would be overestimated.

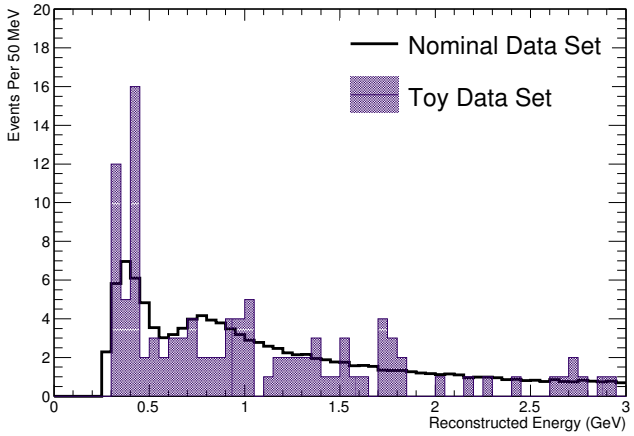


Figure 7.1.: Nominal data set compared to a toy data set for the SK $1R_\mu$ selection.

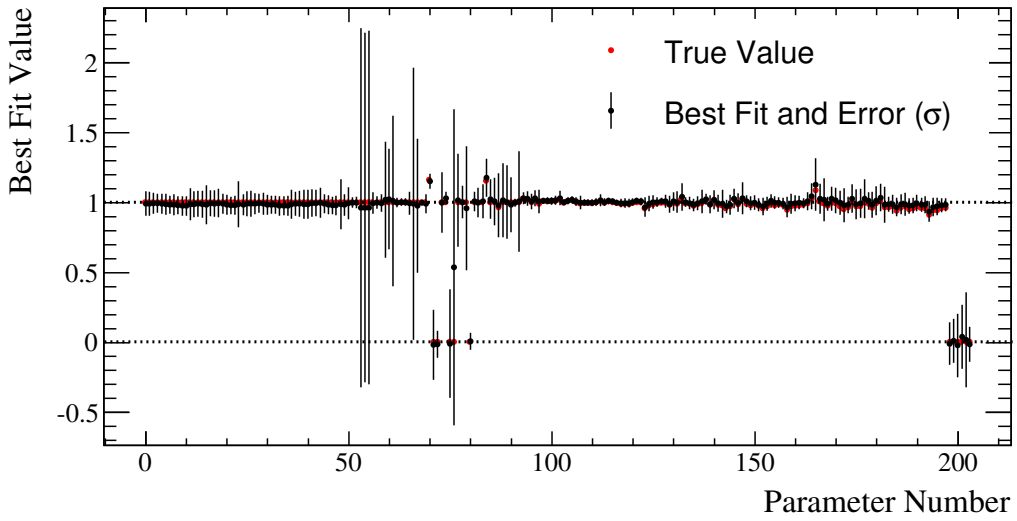
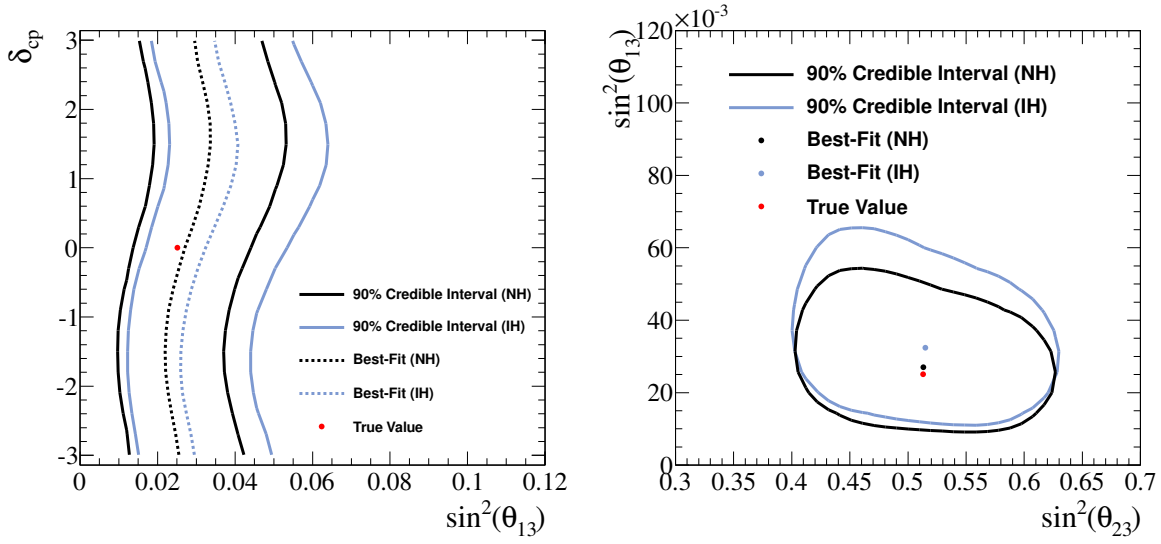


Figure 7.2.: Nominal best fit values from systematic parameters. Error bars are the posterior error in units of σ. Most parameters have either a true central value of 1 or 0. The anomalous point at parameter number 76 is the SK spectral function, which is non Gaussian. Parameter names can be found in Appendix A.



(a) **(b)**

Figure 7.3.: Nominal data fit contours, showing the contours for P(NH)=1 (black) and P(IH)=1 (blue). In (a), best fit lines are constructed in slices of δ_{cp}, and the value at $\delta_{cp} = 0$ is positively offset from the true value due to marginalization of the spectral function as shown in Figure 7.9.

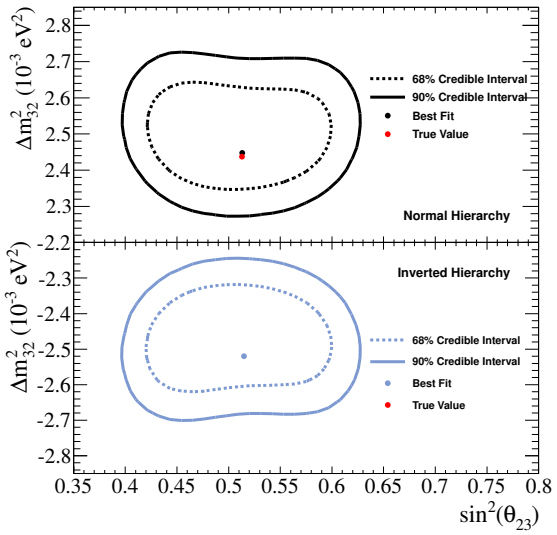


Figure 7.4.: $\sin^2(\theta_{23})$ vs Δm^2_{32} constructed for P(NH)=1 (black) and P(IH)=1 (blue) for a nominal data fit.

The thrown central values of systematic parameters are introduced in the multivariate Gaussian constraint term of the likelihood, as shown in Equation 7.1:

$$-\ln P = \sum_{i=0}^{n}\sum_{j=0}^{n}\frac{1}{2}(p_i^{prop} - p_i^{nom})V_{ij}^{-1}(p_j^{prop} - p_j^{nom}) \tag{7.1}$$

where p^{nom} is the central value which is thrown separately for each toy dataset, according to the prior PDF for that systematic, including the correlation between related systematics, and p_i^{prop} is the value of the parameter at the point of evaluation for the likelihood. Toy experiments are fit using a minimum of 10^6 steps to allow the production of many chains, whilst ensuring adequate convergence.

The toy experiments are used to test the analysis for bias and correct error determination by calculating the pull for each parameter. For parameters that satisfy the condition $\sigma_{prior} < \sigma_{fit}$, the following pull definition is used:

$$pull = \frac{\mu_{fit} - \mu_{true}}{\sigma_{fit}}. \tag{7.2}$$

For all other parameters, the nominal thrown value μ_{throw} and fit error σ_{fit} are incorporated:

$$pull = \frac{\mu_{fit} - \mu_{throw}}{\sqrt{\sigma^2_{prior} - \sigma^2_{fit}}}, \tag{7.3}$$

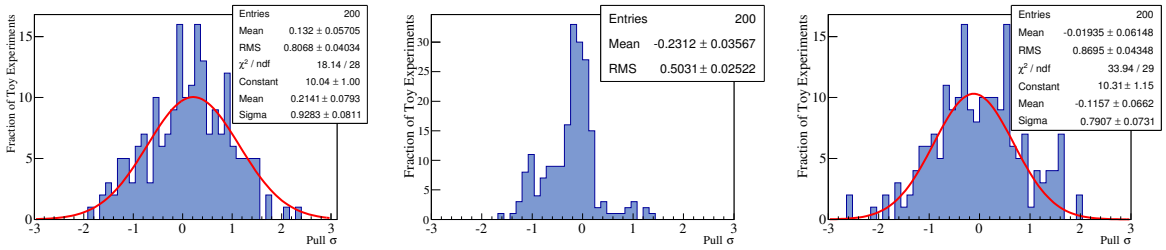
which takes into account the correlated constraint on the parameter.

The best fit point and posterior error for nuisance parameters are extracted from each toy posterior distribution by constructing a 1D marginal distribution for each parameter and fitting a Gaussian to a restricted range defined by mean $\pm$ RMS of the histogram. For oscillation parameter pulls, the best fit is defined as the 3D posterior mode at $\delta_{cp} = 0$, which is determined using the KDE method described in Section 4.2.1. Because the 1D posterior distributions for the oscillation parameters are non-Gaussian, the RMS is used as a better estimate of the error.

The posterior error σ_{fit} of each parameter for every toy experiment was plotted against the prior error and, where available, the ND280 BANFF post-fit error value [106] in Figure 7.6. There is good agreement with the constraining power of the BANFF analysis, and any differences are due to the presence of the SK data term in the likelihood in this analysis, and also the fact that for this analysis, non-Gaussian uncertainties can affect the most probable value in a 1D marginal distribution. Parameters 3 and 7 of the cross-section parameters correspond to the ND280 and SK spectral function parameters respectively. These parameters have a flat prior, and thus have no encouragement to form a Gaussian-like probability distribution. As a result the information about these parameters in the pull plots is of little value because they are not expected to produce a standard pull distribution.

The plots in Figure 7.5 show the oscillation parameter pull distributions, with Gaussian fit results displayed along with the histogram mean and RMS. For $\sin^2(\theta_{13})$ and Δm^2_{32}, the mean and RMS are reasonable, however the pull distribution for $\sin^2(\theta_{23})$ is visibly distorted. The reason for this is that T2K has very little sensitivity to the δ_{cp} phase, and so the distribution of probability in the δ_{cp} parameter causes a shift in $\sin^2(\theta_{23})$. When δ_{cp} is fixed, the pull distribution behaves as expected.

Shown in Figure 7.7 is the mean and RMS of the pull distributions for the toys, constructed from Equations 7.2 and 7.3. In general there is negligible bias in the systematic parameters, but there is a noticeable apparent anti-correlation between the flux and ND280 detector errors. This is because the parameter sets are degenerate; they



(a) $\sin^2(\theta_{13})$ pull distribution **(b)** $\sin^2(\theta_{23})$ pull distribution **(c)** Δm^2_{32} pull distribution

Figure 7.5.: Oscillation parameter pulls for $\delta_{cp} = 0$. Since there is no sensitivity to fit for δ_{cp}, the pull distribution is omitted.

are both normalization parameters for the same kinematic variables. This, combined with the fact that they are highly correlated not only with themselves but also with the SK flux parameters, supports this conclusion. When toy experiments are generated with the ND280 detector errors fixed at nominal values, the flux bias is removed, as shown in Figure 7.8.

7.1.3. Marginalization Induced Biases

To extract the best fit values and errors necessary for pull calculations, the 1D marginal posterior for each parameter is constructed as described in Section 7.1.2. This method means that for each parameter, the best fit estimate and error is found by marginalizing all other parameters. In doing so, any non-Gaussian behaviour and correlations with parameters with non-Gaussian behaviour can cause apparent biases in the mean of the pull distributions. In this analysis, there are several parameter pulls which are not within 1σ of 0. These are:

- Quasi-Elastic Axial Mass (M_A^{QE}): this parameter is correlated with the ND280 spectral function parameter, which is both non-Gaussian and one-sided (Figure 7.10).
- Fermi momentum: this parameter is highly correlated with spectral function, which likewise produces an apparent shift as with M_A^{QE}. (Figure 7.11).
- Spectral function for carbon and oxygen: these are parameters which are defined to be between 0 and 1, and have a distinctly non-Gaussian shape in the posteriors.
- CCnQE ν_μ Normalization: this parameter is correlated with the oscillation parameters. Since these parameters have non-Gaussian posterior distributions, marginal-

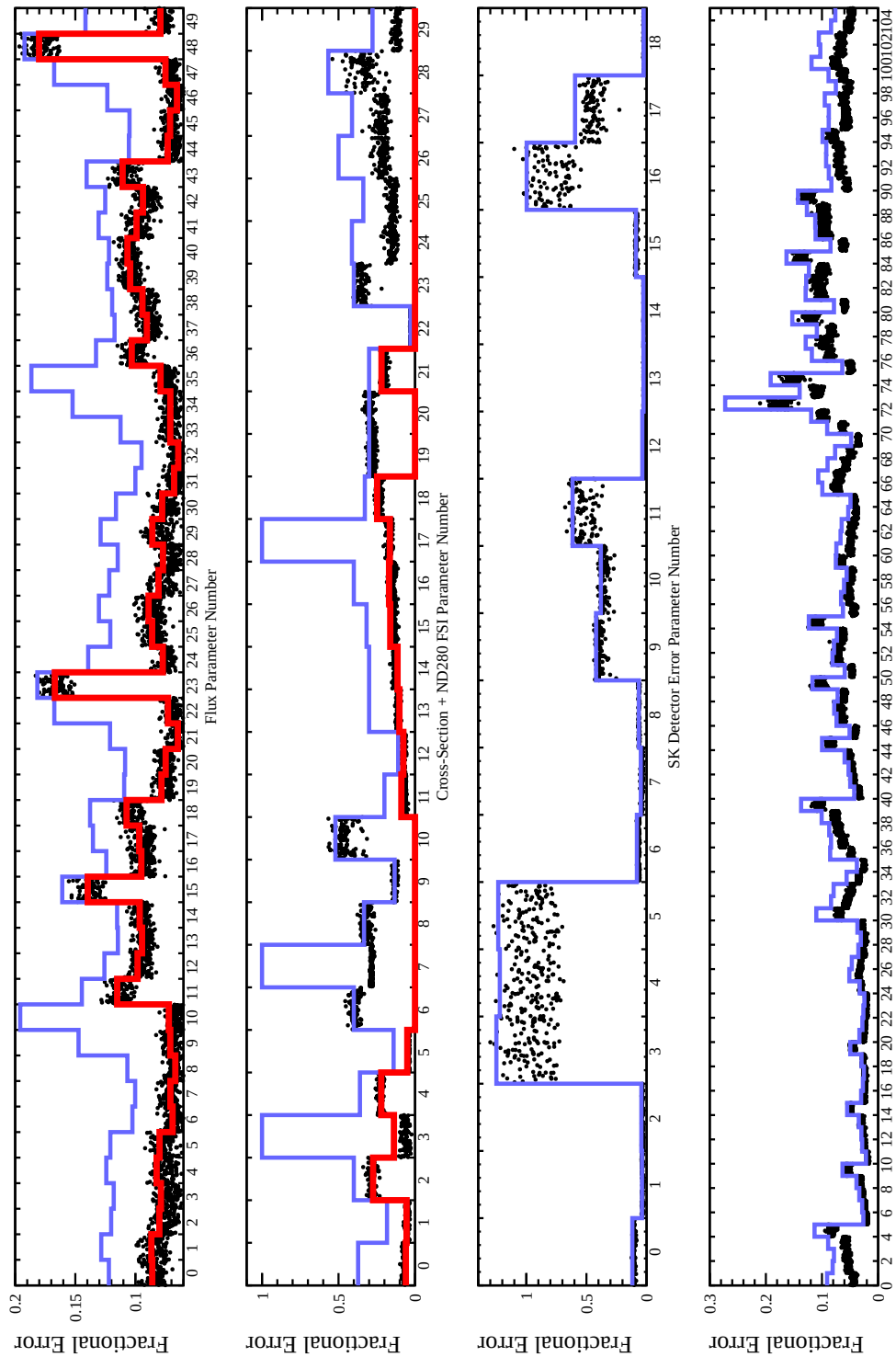


Figure 7.6.: Blue line: prior error on parameter. Red line: BANFF post-fit error (where applicable). Black points: posterior error from toy experiments. These plots show how the power to constrain parameter errors is in good agreement with the BANFF post-fit values. Parameter names can be found in Appendix A.

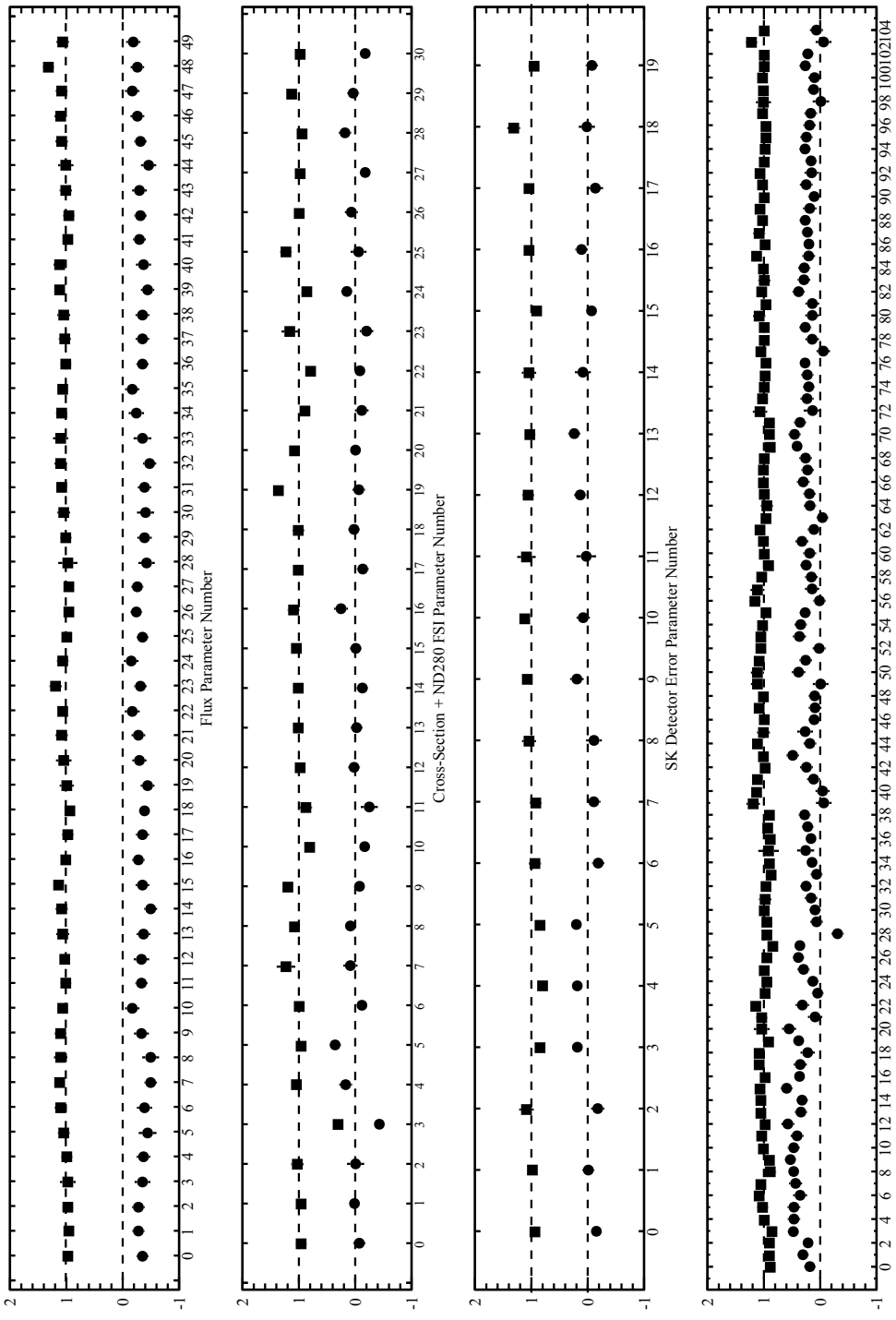
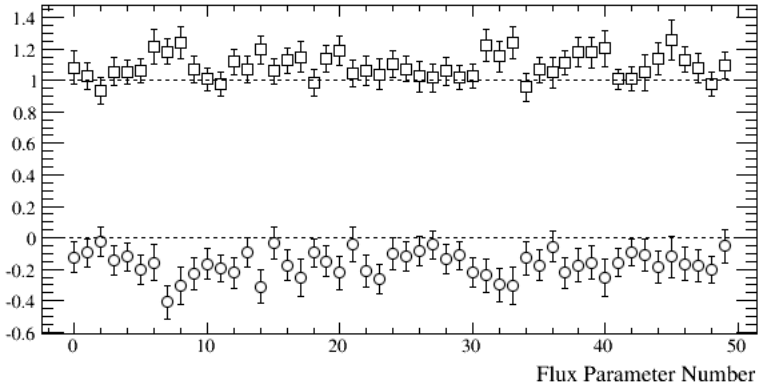
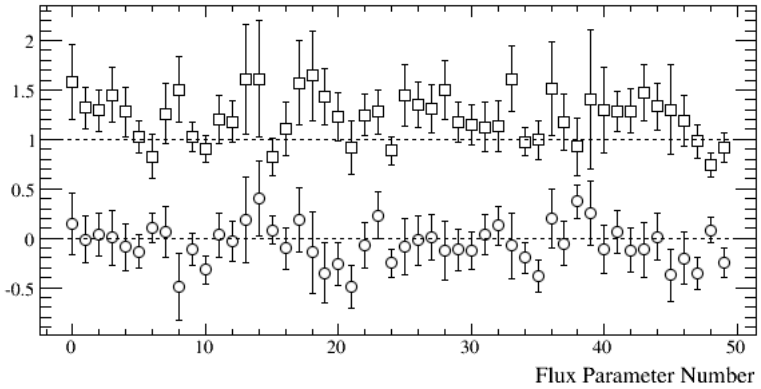


Figure 7.7.: Mean and RMS of the pull distributions for each systematic parameter. Parameter names can be found in Appendix A.



(a) All parameters free. 100 toy experiments.



(b) ND280 Detector parameters fixed at nominal. 50 toy experiments.

Figure 7.8.: The mean and RMS of the flux parameter pulls, for the case of all parameters free 7.8a and also for the case where all ND280 detector error parameters are fixed at their nominal values 7.8b. When the ND280 detector errors are fixed, the bias in the flux parameters is not present.

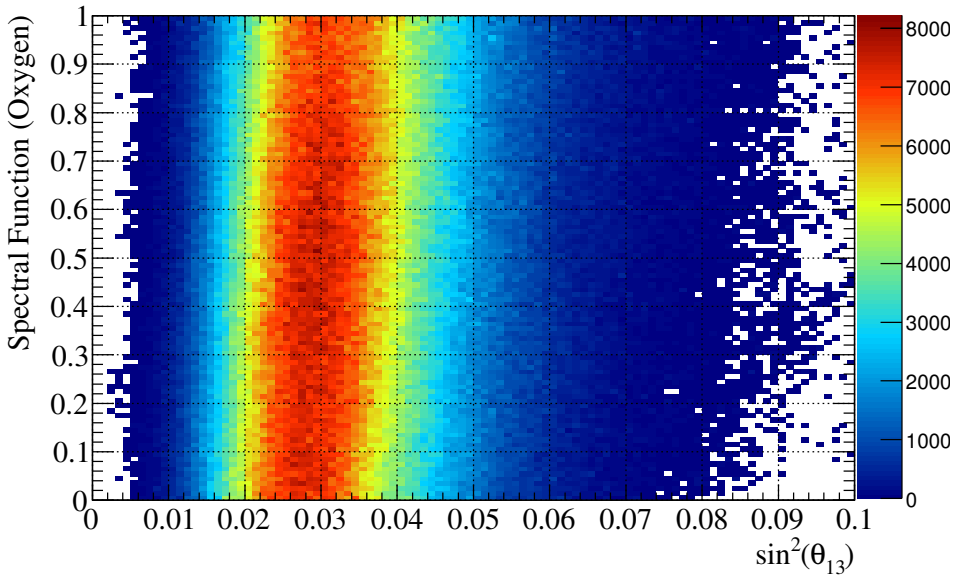


Figure 7.9.: Joint posterior for $\sin^2(\theta_{13})$ and the oxygen spectral function. When marginalizing the spectral function, due to the correlations between both parameters and the boundary at 0, a shift in probability to positive values is caused in the 1D marginal posterior of $\sin^2(\theta_{13})$. Plot constructed from a nominal data set posterior.

izing them affects the posteriors of correlated parameters. This manifests in the CCnQE ν_μ normalization parameter as a small negative shift in the central value.

Although the pull distributions appear biased, the bias only enters due to the method of extracting the best fit value. The marginalization into 1D causes a bias to appear because of the non-Gaussian behaviour of a parameter in multiple dimensions, or because there is a boundary in the parameter space. Because all the best fit values for the pulls were calculated in 1D space, they are susceptible to this biasing effect. However, as the bias is due to the estimation method, rather than the analysis mechanism itself, it can safely be ignored, knowing that the most dense region yields the expected value. The most dense region must be found in multiple dimensions using a density estimation technique, as discussed in 4.2.1, however for the purpose of pulls, the simpler method of finding the mean of the posterior distribution is adequate.

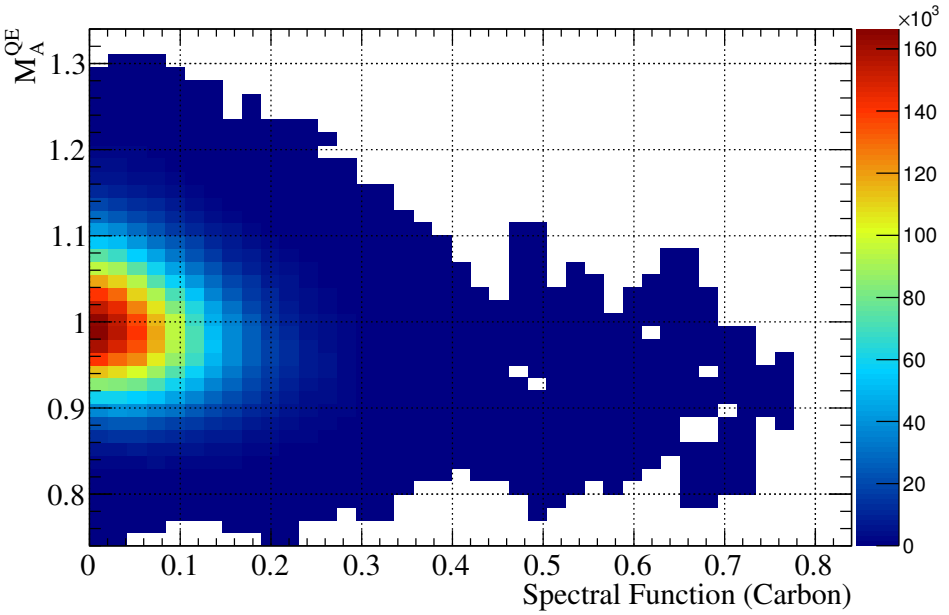


Figure 7.10.: Correlation between quasi-elastic axial mass and spectral function parameters for carbon. Plot constructed from a nominal data set posterior.

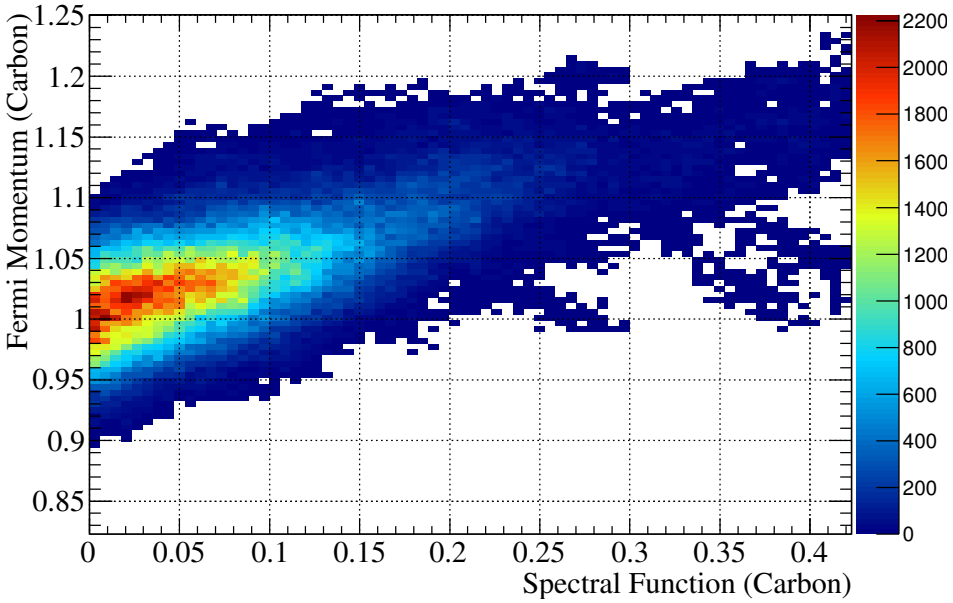


Figure 7.11.: Correlation between fermi momentum and spectral function parameters for carbon. Plot constructed from a nominal data set posterior.

7.1.4. Mock Data Set Fits

A series of 6 "fake" data sets (FDS) were produced. The parameter values used to generate these data sets are denoted in Table 7.1. Best-fit points for T2K only fits are found using the adaptive KDE method with δ_{cp} fixed at the true value. This is because the FDS are generated with no reactor constraint, and so there is little sensitivity to δ_{cp}.

Table 7.1.: Table showing the configuration of the fake data sets. **Bold** elements highlight the defining parameter value of that data set.

Fake Data Set	Mass Hierarchy	$\sin^2(\theta_{23})$	Δm^2_{32}	$\sin^2(\theta_{13})$	δ_{cp}	Systematic
0	NH	0.513	2.4375	0.0251	0	Nominal
1	NH	**0.37**	2.4375	0.0251	0	Random Throws
2	NH	0.513	**2.75**	0.0251	0	Random Throws
3	NH	0.513	2.4375	**0.04**	0	Random Throws
4	NH	0.513	2.4375	0.0251	**-π/2**	Nominal
5	**IH**	0.513	2.4375	0.0251	0	Nominal

Table 7.2 shows the results of the analysis when using the FDS. There is generally good agreement between the results of the analyses and the true values. The value of $\sin^2\theta_{13}$ shows the most discrepancy, and this is due to the marginalization of the spectral function as shown in Figure 7.9. There is an interesting feature of FDS 1, where the input value was an off-maximal value of $\sin^2\theta_{23} = 0.37$. A profiling technique used in a frequentist analysis would find the best fit value of $\sin^2\theta_{23}$ to be in the upper octant, whereas this analysis finds the most probable region to be in the lower octant. This is demonstrated in Figure 7.12, where the posterior is restricted to a smaller region around the best fit points in $\sin^2\theta_{13}$ and Δm^2_{32}—a technique similar to the profiling method of the minimizer fit— and there is greater posterior density found in the upper octant.

7.1.5. Comparison with BANFF Matrix Fit

The analysis framework is modular and so can also be configured to constrain the SK flux and cross-section uncertainties using the BANFF matrix instead of using the ND280 data directly. This is performed only in this section, and for the purpose of validation.

A comparison of the contours and best fit points (Figure 7.13) produced with both methods when fitting FDS 5 was made and the results show negligible difference between

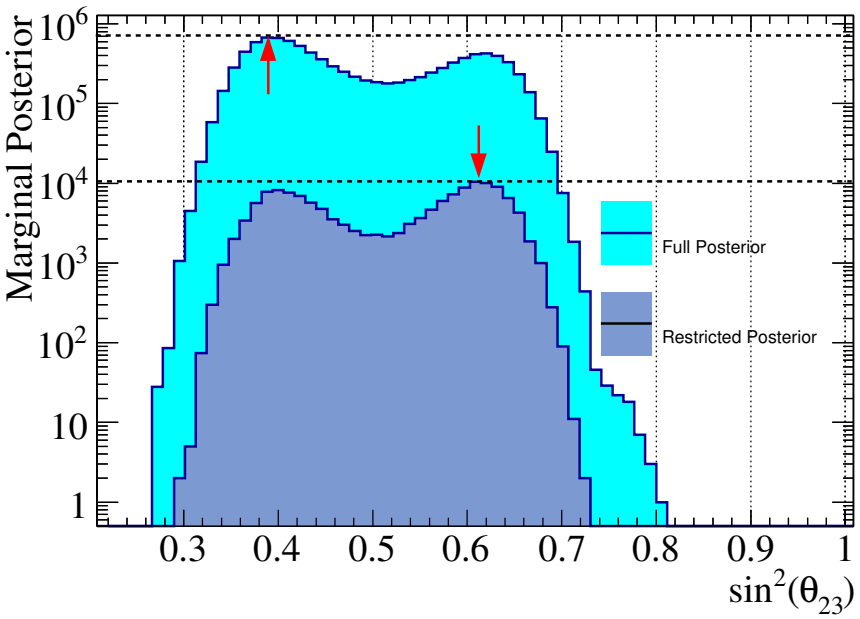


Figure 7.12.: The full marginal posterior of $\sin^2(\theta_{23})$ for FDS 1 (cyan) compared with a "restricted posterior" constructed from MCMC steps taken only from a small region around the best fit point of $\sin^2(\theta_{13})$ and Δm^2_{32} (darker blue). Restricting the posterior to points only around the most probable regions of the marginalized oscillation parameters is similar in approach to the frequentist profiling technique. Red arrows indicate the 1D posterior mode for each distribution. This exercise highlights the difference between analysis techniques.

Table 7.2.: Normal hierarchy best-fit comparison table between analysis results and the true values for all fake data sets with no reactor constraint.

FDS	$\Delta m^2_{32} \times 10^{-3}$	$\sin^2(\theta_{23})$	$\sin^2(\theta_{13})$	δ_{cp}
0 True	**2.400**	**0.513**	**0.0251**	**0**
0	2.380	0.521	0.0399	0 (fixed)
1 True	**2.400**	**0.37**	**0.0251**	**0**
1	2.228	0.409	0.0181	0 (fixed)
2 True	**2.713**	**0.513**	**0.0251**	**0**
2	2.562	0.509	0.0202	0 (fixed)
3 True	**2.400**	**0.513**	**0.04**	**0**
3	2.528	0.533	0.0558	0 (fixed)
4 True	**2.400**	**0.513**	**0.0251**	**-π/2**
4	2.415	0.526	0.462	-π/2 (fixed)
5 True	**-2.400**	**0.513**	**0.0251**	**0**
5	-2.571	0.511	0.0230	0 (fixed)

the two results. This implies that the multivariate Gaussian constraint from ND280 used in previous T2K oscillation results is an adequate approximation with current statistics, however this is not expected to hold true as more POT is collected. By avoiding the Gaussian approximation of the ND280 constraint, this analysis serves as a validation of the BANFF matrix, and also as a proof-of-concept for future analyses.

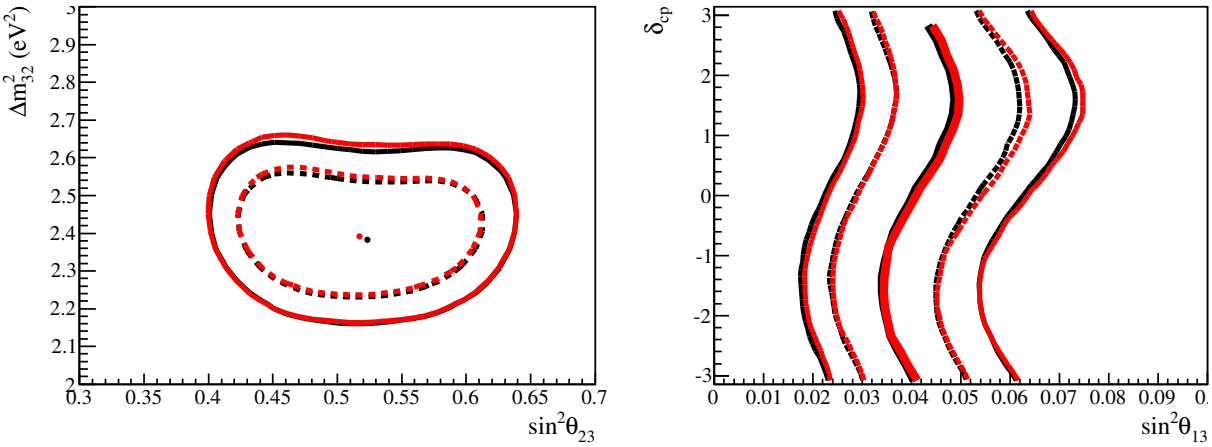


Figure 7.13.: The comparison of contours and best fit points between the BANFF extrapolation (red) and the simultaneous fit with ND280 data methods. The contours suggest that the simultaneous fitting method yields a smaller uncertainty, however the difference is considered negligible. These comparisons have been done with the mass prior P(NH)=1.

7.2. Data Analysis Results

This section details the results obtained from analysing simultaneously the SK and ND280 Run 1–4 data, totalling 6.57×10^{20} and 5.90×10^{20} protons on target respectively. For these analyses, the solar sector oscillation parameters have Gaussian priors of $\sin^2\theta_{12} = 0.311 \pm 0.017$ and $\Delta m^2_{21} = 7.5 \pm 0.2 \times 10^{-5}$ eV^2. For the oscillation probability reweighting calculations, an average earth density of 2.6 g cm^{-3} [107] is used, for a baseline of 295 km [35]. The earth density and baseline are held constant.

The analysis assumes the atmospheric mass-squared splitting to be Δm^2_{32} in both MH, meaning that they correspond to the same difference between mass eigenstates, but have a different splitting size as seen in Figure 1.13. This convention is chosen because it reflects more naturally what T2K has the ability to measure.

These results are reported with different priors on the mass hierarchy (called the *mass prior*). One choice of prior is to choose P(NH)=1 or P(IH)=1. The other choice is to choose P(NH)=P(IH)=0.5. The P(NH)=1 or P(IH)=1 mass prior is the equivalent of doing an analysis assuming NH or IH. This implies that the analysis must be repeated with both NH or IH assumed, which will result in two posterior distributions which are both normalized to 1. This means that the relative probability between the NH and IH states is lost. The P(NH)=P(IH)=0.5 mass prior retains the relative probability between the two MH states, and so will not only give more information about which MH is preferred by the data, but also produces one set of credible intervals which are marginalized over the MH in the case when Δm^2_{32} is not part of the marginal posterior i.e. when Δm^2_{32} has been marginalized. This simplifies the presentation of results, as one can report a single credible region in ignorance of whichever MH is hypothesis is true.

7.2.1. T2K Run 1–4 Data Analysis

The was first performed using only T2K data, with a Markov chain of 1.8×10^7 steps after burn-in. For this analysis, as there is little constraint on δ_{cp}, the best fit point is found by fixing δ_{cp} at 21 sequential steps in the range of $-\pi$ to π, and fixing the δ_{cp} parameter in the 4D adaptive kernel estimation to find the best fit in 3D for the other oscillation parameters. The result is a line of most probable values of δ_{cp}.

Table 7.3.: Best-fit values for oscillation parameters extracted from the marginal posterior of the Run 1–4 data.

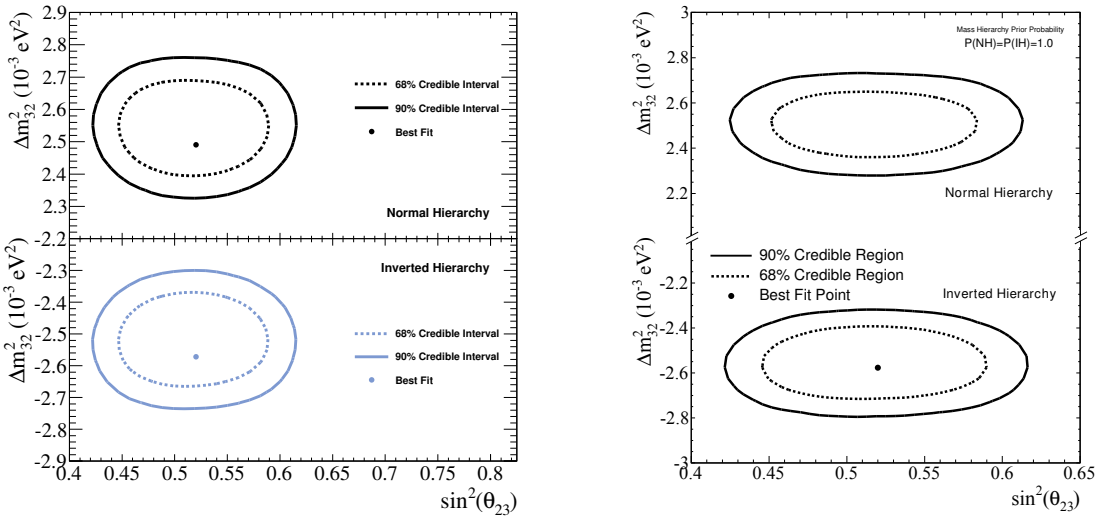
	$\lvert\Delta m^2_{32}\rvert$	$\sin^2(\theta_{23})$	$\sin^2(\theta_{13})$	δ_{cp}
Normal Hierarchy	2.491	0.520	0.0377	0 (fixed)
Inverted Hierarchy	2.571	0.520	0.0454	0 (fixed)

Table 7.3 shows the best fit parameters in the $\delta_{cp} = 0$ slice. This slice is chosen rather arbitrarily, but it aids in comparison to other measurements which report results with $\delta_{cp} = 0$. Credible regions are produced in 2D for several different sets of parameters; these contours are produced marginalized over all other parameters, but constructed separately for the choice of mass prior of P(NH)=1 or P(IH)=1; or produced with the choice of mass prior P(NH)=P(IH)=0.5. Figure 7.14 shows the contours in $\sin^2(\theta_{23})$–Δm^2_{32} space. When comparing to Figure 7.4, the contours in Figure 7.14 are seen to be smaller than the nominal dataset. This is because T2K saw fewer ν_μ disappearance events at SK than expected, and prefers maximal values of θ_{23}. When looking at Figure 6.17, the minimum number of events (z axis) for maximal mixing is higher than the number of data events collected, and so the probability is constrained to near the physical boundary of maximal mixing, creating smaller contours.

Figure 7.15 shows the contours in $\sin^2(\theta_{13})$–δ_{cp} space, where the best fit is shown as a line connecting the best fit values in the slices of δ_{cp}. Figure 7.16 shows the contours in $\sin^2(\theta_{23})$–$\sin^2(\theta_{13})$.

Figure 7.17 shows the 1D credible intervals for $\sin^2(\theta_{13})$, $\sin^2(\theta_{23})$, and Δm^2_{32}, where all other parameters are marginalized. In this case, the intervals are shown for P(NH)=1 in red, P(IH)=1 in blue, and P(NH)=P(IH)=0.5 in black.

Figure 7.18 shows the best fit spectra of the Run 1–4 SK data constrained by the ND280 data, for $1R_\mu$ and $1R_e$ samples. The best fit spectra is determined via a marginalization method. This method is as follows: 2500 random steps are drawn ("sampled") from the MCMC, and with each sample the parameter values are used to calculate the expected event rate per bin of the energy spectra; this is essentially marginalizing over all parameters, oscillation included, to find the posterior distribution in each energy bin. The combination of all the samples creates a distribution of event rates for each bin. Finally, for each bin, a Gaussian is fitted around the peak of the event rate distribution, and the mean of the fit is taken to be the predicted value for that bin. Most bins take on a Gaussian shape, but in some bins, especially near the oscillation



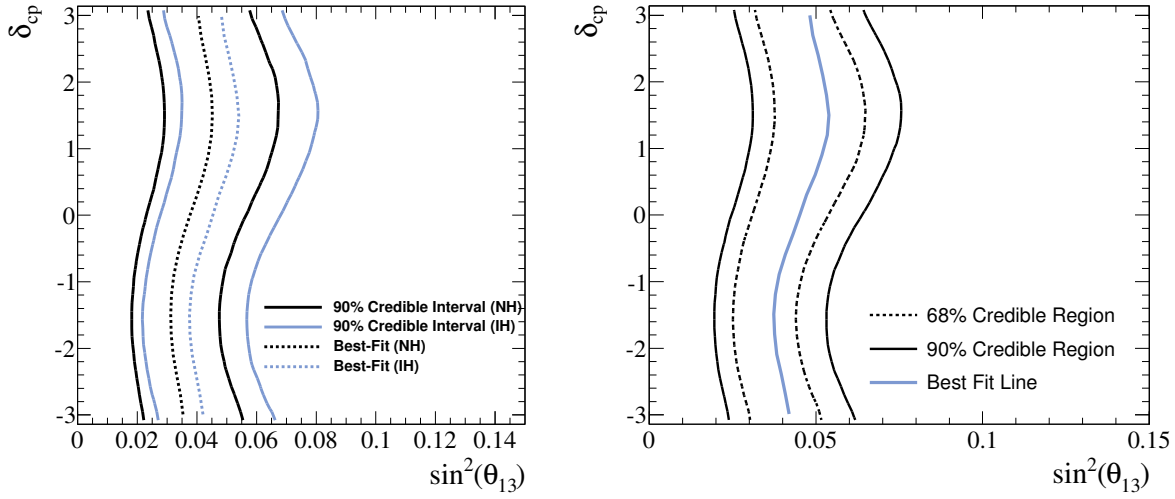
(a) **(b)**

Figure 7.14.: Run 1–4 data fit 2D contours in Δm^2_{23}–$\sin^2(\theta_{23})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit points are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5. The best fit point shown is the IH best fit point, which is the preferred hierarchy for the T2K-only fit.

maxima in the 1R$_\mu$ sample, the distribution is non-Gaussian, due to the influence of the nearby physical boundary in $\sin^2\theta_{23}$.

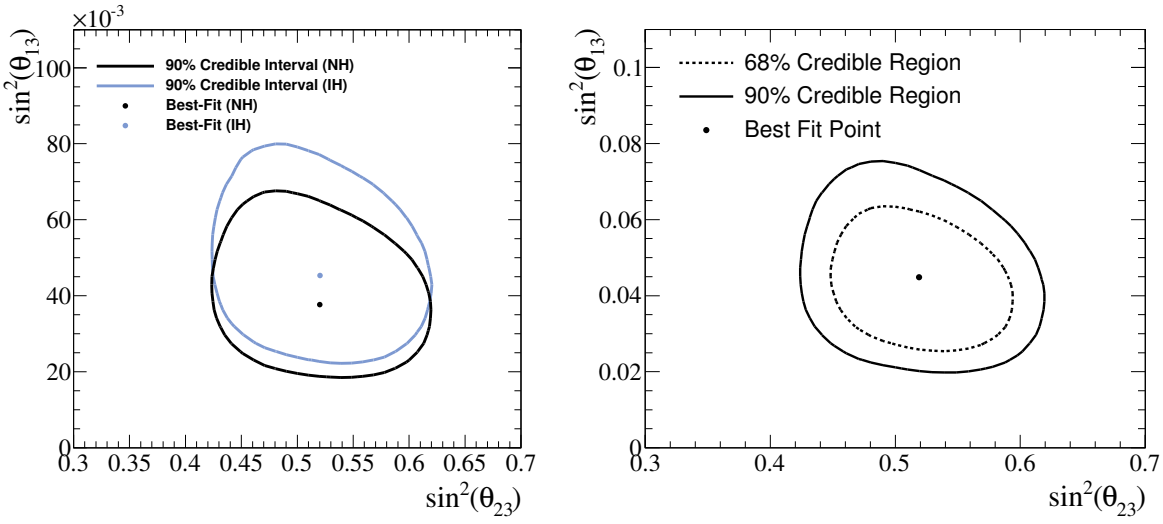
Figure 7.19 shows the best fit values of all systematic parameters, along with their posterior error (in units of σ) and the central values used for the constraint term. Parameter number corresponds to M_A^{RES}, and its most likely value appears lower than the prior value. This is because the prior value determined by NIWG from fits to MiniBooNE data [101] (1.41 GeV) was considered unrealistically high, and in fact the postfit value (approximately 1 GeV) is more in line with theoretical expectations.

Figure 7.20 shows the momentum and angle distributions for ND280 with the pre-fit MC prediction and post-fit spectra, calculated in the same way as for the SK spectra.



(a) **(b)**

Figure 7.15.: Run 1–4 data fit 2D contours in δ_{CP}–$\sin^2(\theta_{13})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit lines are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5, and marginalizing over the mass hierarchy. The best fit line shown is the IH best fit line, which is the preferred hierarchy for the T2K-only fit.



(a) (b)

Figure 7.16.: Run 1–4 data fit 2D contours in $\sin^2(\theta_{23})$–$\sin^2(\theta_{13})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit points are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5, and marginalizing over the mass hierarchy. The best fit point shown is the IH best fit point, which is the preferred hierarchy for the T2K-only fit.

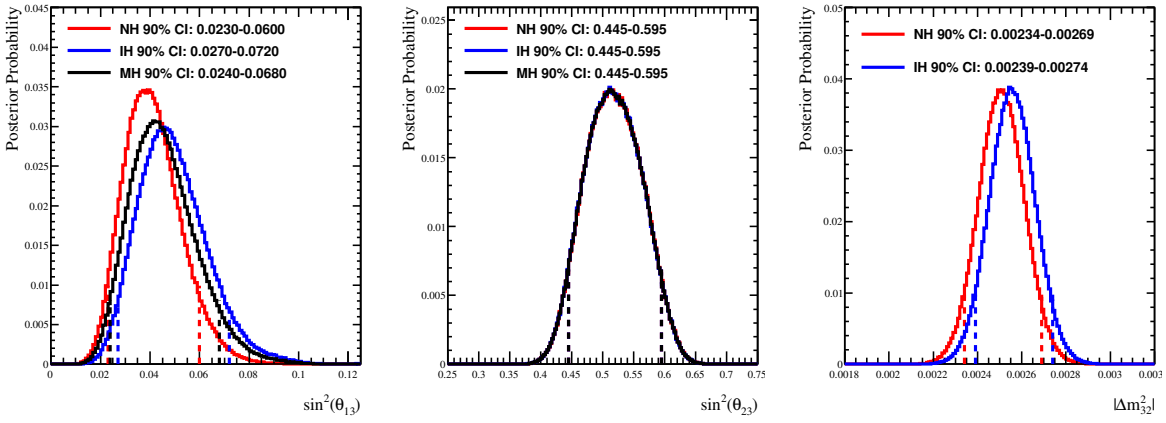


Figure 7.17.: Credible intervals in 1D for $\sin^2(\theta_{13})$, $\sin^2(\theta_{23})$, and $|\Delta m^2_{32}|$. The PDFs for the angles are shown for the mass prior of P(NH)=1 in red, the mass prior P(IH)=1 in blue, and the mass prior P(NH)=P(IH)=0.5 in black, marginalized over the hierarchies. The PDF for the mass splitting is shown only for the mass prior of P(NH)=1 in red and the mass prior P(IH)=1 in blue. The 90% credible intervals are shown by the dotted lines and given in the plot legends.

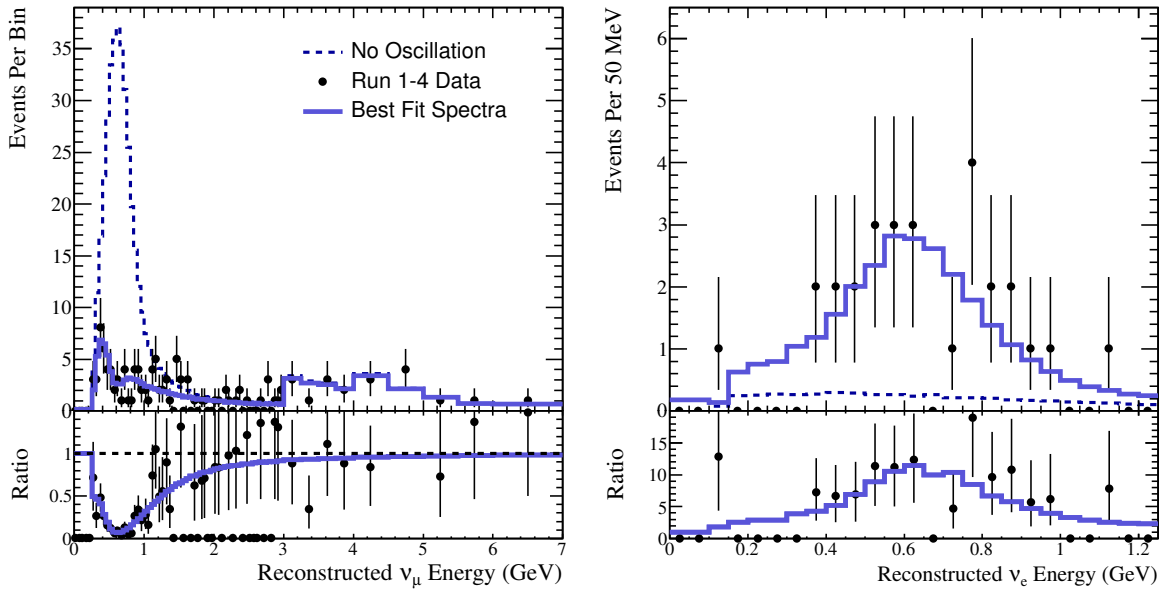


Figure 7.18.: Run 1–4 data best fit spectra for SK $1R_\mu$ and $1R_e$ samples.

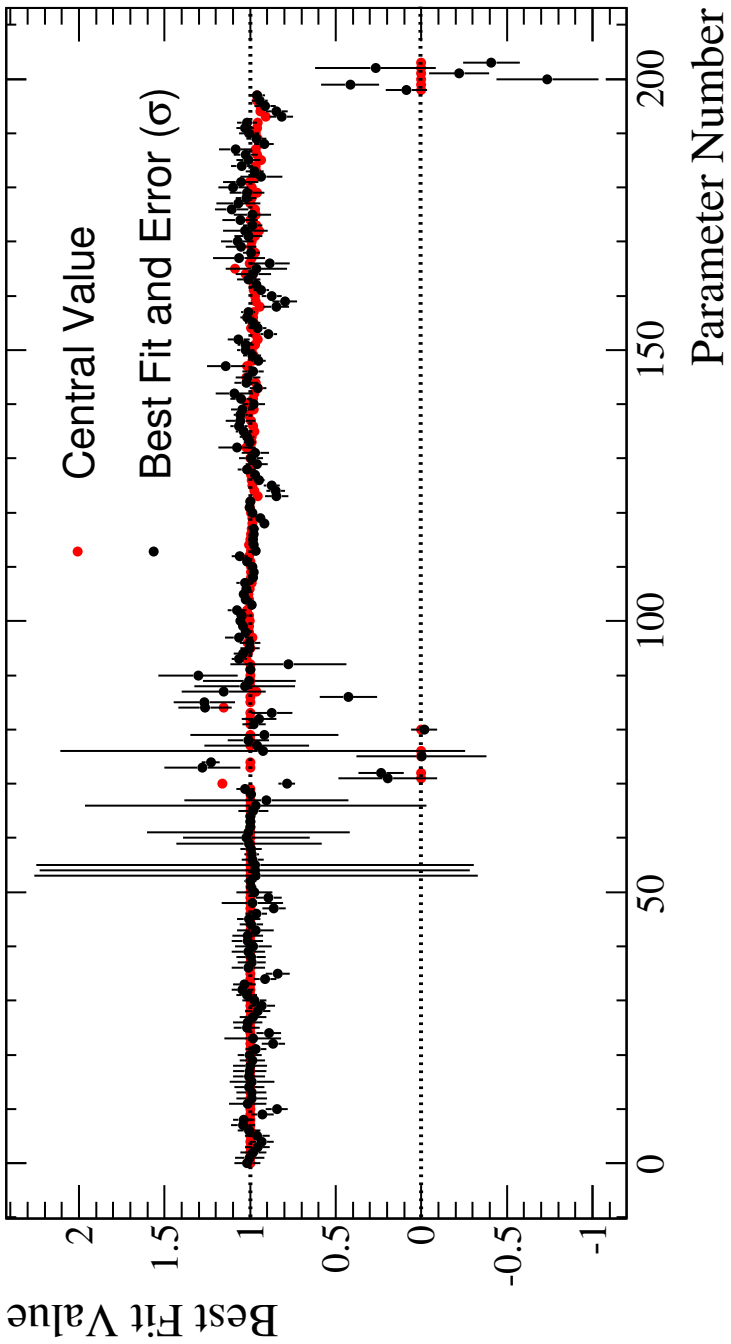


Figure 7.19.: Best fit values of the systematic parameters for the T2K only data fit. Shown also is the central value used in the constraint term. Parameter names can be found in Appendix A.

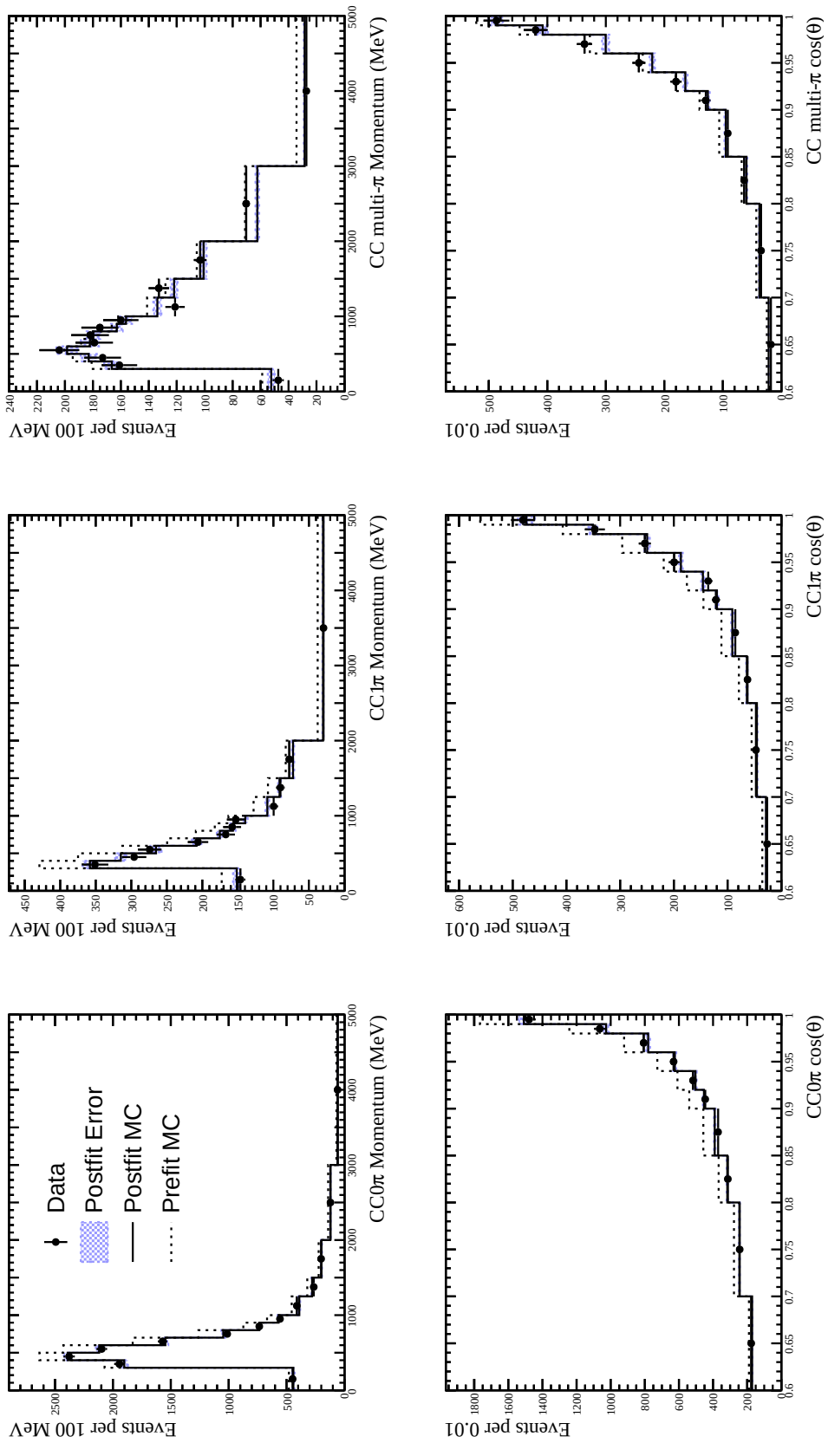


Figure 7.20.: Comparison of pre- and postfit MC prediction with data, projected into momentum (top row) and the $\cos\theta$ (bottom row) axes. Shown left-to-right are the CC0π sample, the CC1π sample, and the CCother sample. Errors are determined using the same method as for the best fit spectra (Figure 7.18) and using the standard deviation of each bin as the posterior error.

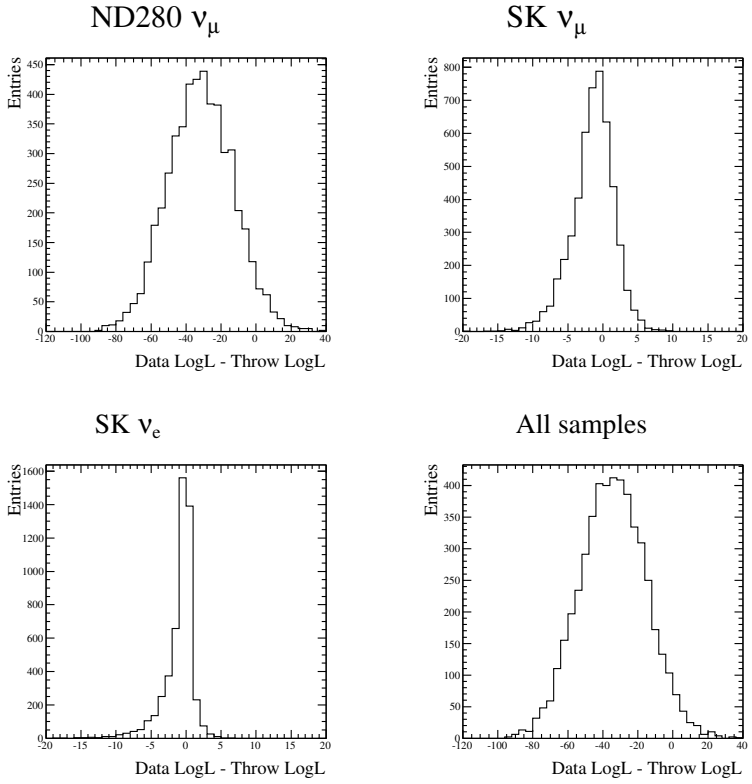


Figure 7.21.: Goodness-of-fit distributions for the three different samples in the fit and the summed total. The p-value is the percentage of each distribution which is greater than zero.

A goodness-of-fit is calculated as in [108], where at each chain sample used for the best fit spectra, a fake dataset is thrown from the MC prediction for that sample. The log likelihood ratio between the fake dataset and the MC prediction is calculated, as is the log likelihood ratio between the real data and the MC prediction. A p-value is calculated as the percentage of samples for which the data better fit the MC prediction than the fake data. In order to have $N > 10$ in each bin, a requirement for this method, the $1R_\mu$ sample is rebinned into five bins (0–0.4; 0.4–0.7; 0.7–1.0; 1.0–2.0; and 2.0–30.0 GeV) and the $1R_e$ sample is considered as one bin only. The ND280 sample is considered in the bins used to fit the data. This means that the overall p-value is completely dominated by the ND280 sample. Figure 7.21 shows the ND280, $1R_\mu$, $1R_e$, and total distributions for the quantity $\ln L_{data} - \ln L_{throw}$; the p-value is the percentage of this distribution above zero. The p-values are: ND280-only, 0.044; SK $1R_e$, 0.32; SK $1R_\mu$, 0.35; and all samples, 0.036. These values indicate no disagreement with data for the SK samples. The value for the ND280 samples is somewhat low, indicating some disagreement; however, this is a known effect [106], and the agreement between the results of the ND280 fits for both

MaCh3 and the minimizer BANFF fit [32] and the data are nearly equivalent. Studies show that the goodness-of-fit is sensitive to the number of MC events used to construct the ND280 PDFs, and so it is understood that the relatively low (when compared to SK) number of MC events used in this analysis is causing the disagreement.

Table 7.4.: Best-fit values for oscillation parameters extracted from the marginal posterior of the Run 1-4 data fit with reactor constraint.

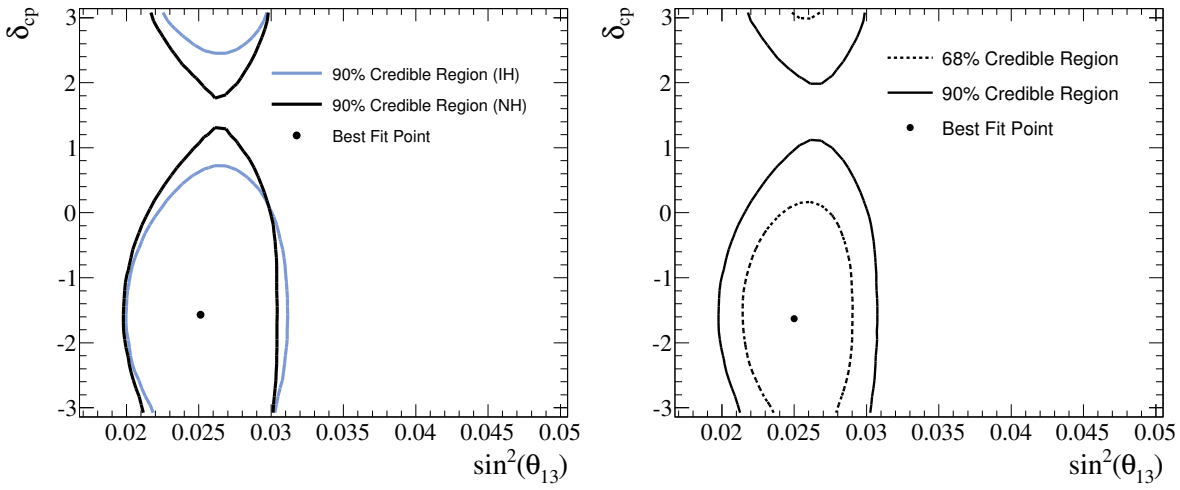
	$\lvert\Delta m^2_{32}\rvert$	$\sin^2(\theta_{23})$	$\sin^2(\theta_{13})$	δ_{cp}
Normal Hierarchy (mass prior P(NH)=1)	2.509	0.528	0.0250	-1.601
Inverted Hierarchy (mass prior P(IH)=1)	2.556	0.533	0.0255	-1.587

7.2.2. T2K Run 1–4 Data Analysis With Reactor Prior on θ_{13}

The data samples were then analysed using T2K data in combination with the PDG 2013 reactor experiment gaussian constraint of $\sin^2(2\theta_{13}) = 0.095 \pm 0.01$, with a Markov chain of 3.168×10^7 steps after burn-in was removed. For this type of analysis, the most probable value is found with a 4D adaptive kernel estimate of the oscillation parameters of interest posterior distribution.

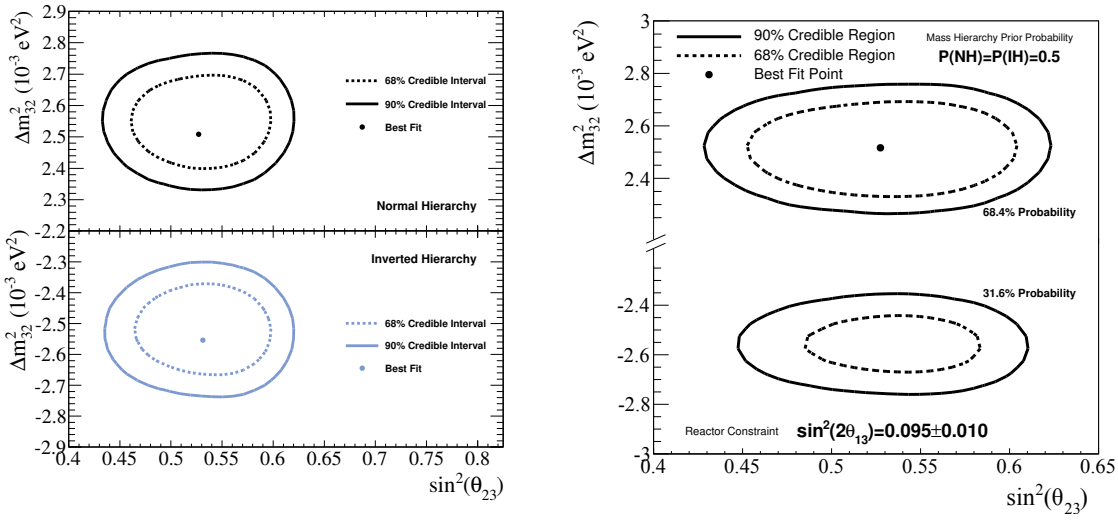
Table 7.4 shows the best fit parameters. Credible regions are produced in 2D for several different sets of parameters; these contours are produced marginalized over all other parameters, but constructed separately for the choice of mass prior of P(NH)=1 (black lines, generally) or P(IH)=1 (blue lines, generally); or for the choice of mass prior P(NH)=P(IH)=0.5.

Figure 7.23 shows the contours in $\sin^2(\theta_{23})$–Δm^2_{32} space. Figure 7.22 shows the contours in $\sin^2(\theta_{13})$–δ_{cp} space. Figure 7.24 shows the contours in $\sin^2(\theta_{23})$–$\sin^2(\theta_{13})$.



(a) (b)

Figure 7.22.: Run 1–4 data fit 2D contours in δ_{CP}–$\sin^2(\theta_{13})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit lines are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5, and marginalizing over the mass hierarchy. The best fit point shown is the NH best fit point, which is the preferred hierarchy for the reactor-constrained fit.



(a) (b)

Figure 7.23.: Run 1–4 data fit 2D contours in Δm^2_{23}–$\sin^2(\theta_{23})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit points are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5. The best fit point shown is the NH best fit point, which is the preferred hierarchy for the reactor-constrained fit.

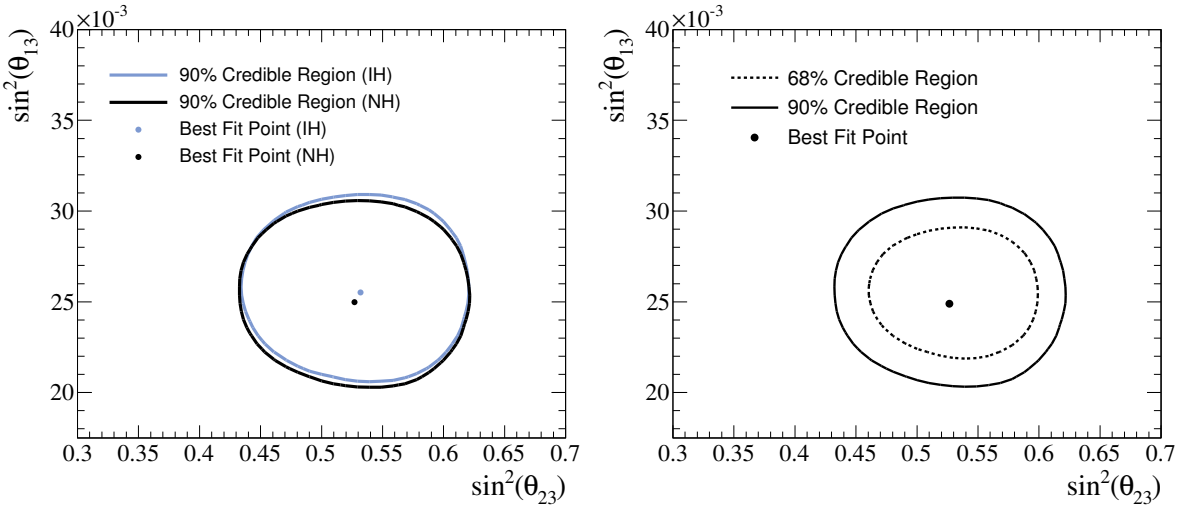


Figure 7.24.: Run 1–4 data fit 2D contours in $\sin^2(\theta_{23})$–$\sin^2(\theta_{13})$ space. The plots in (a) show the contours using the mass prior P(NH)=1 (black) and P(IH)=1 (blue). The respective best fit points are shown. The plots in (b) show the contours using the mass prior P(NH)=P(IH)=0.5, and marginalizing over the mass hierarchy. The best fit point shown is the NH best fit point, which is the preferred hierarchy for the reactor-constrained fit.

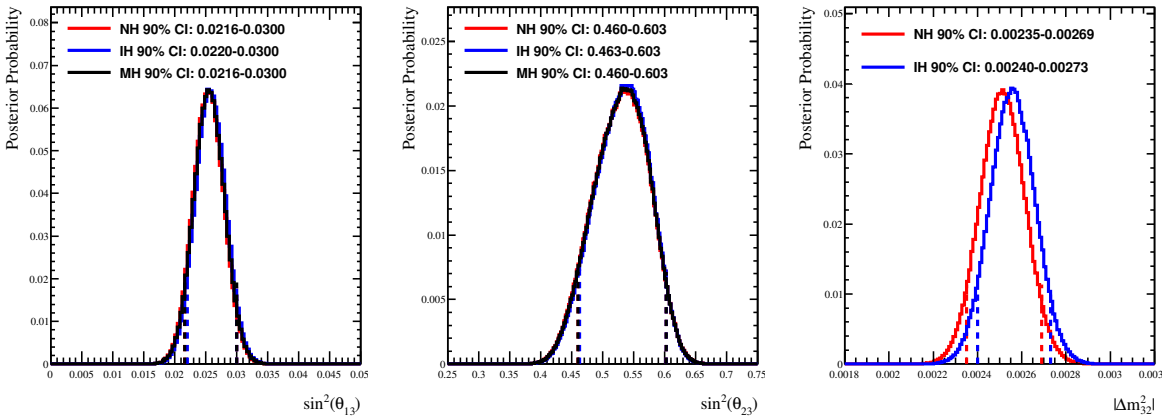


Figure 7.25.: Credible intervals in 1D for $\sin^2(\theta_{13})$, $\sin^2(\theta_{23})$, and $|\Delta m^2_{32}|$. The PDFs for the angles are shown for the mass prior of P(NH)=1 in red, the mass prior P(IH)=1 in blue, and the mass prior P(NH)=P(IH)=0.5 in black, marginalized over the hierarchies. The PDF for the mass splitting is shown only for the mass prior of P(NH)=1 in red and the mass prior P(IH)=1 in blue. The 90% credible intervals are shown by the dotted lines and given in the plot legends.

Sample	Run 1-4 Events	Expectation (T2K Only)	Expectation (Reactor Constraint)
$1R_\mu$	120	123.36	123.93
$1R_e$	28	28.68	22.66

Table 7.5.: Event rate comparison between run 1-4 SK data and the best fit spectra expectations for T2K only and reactor constrained fits.

Figure 7.25 shows the 1D credible intervals for $\sin^2(\theta_{13})$, $\sin^2(\theta_{23})$, and Δm^2_{32}, where all other parameters are marginalized. In this case, the intervals are shown for P(NH)=1 in red, P(IH)=1 in blue, and P(NH)=P(IH)=0.5 in black.

Figure 7.26 shows the best fit spectra for the reactor constrained fit, produced using the same method as for the T2K only fit. Figure 7.27 shows the best fit spectra for the reactor constrained fit compared to the T2K only fit. A comparison of the expected event rates from the best fit spectra is seen in Table 7.5.

Figure 7.28 shows the best fit values of all systematic parameters, along with their posterior error (in units of σ) and the central values used for the constraint term. Once again, the M_A^{RES} posterior fit value is much lower than the prior, for the same reason as discussed in the T2K only analysis.

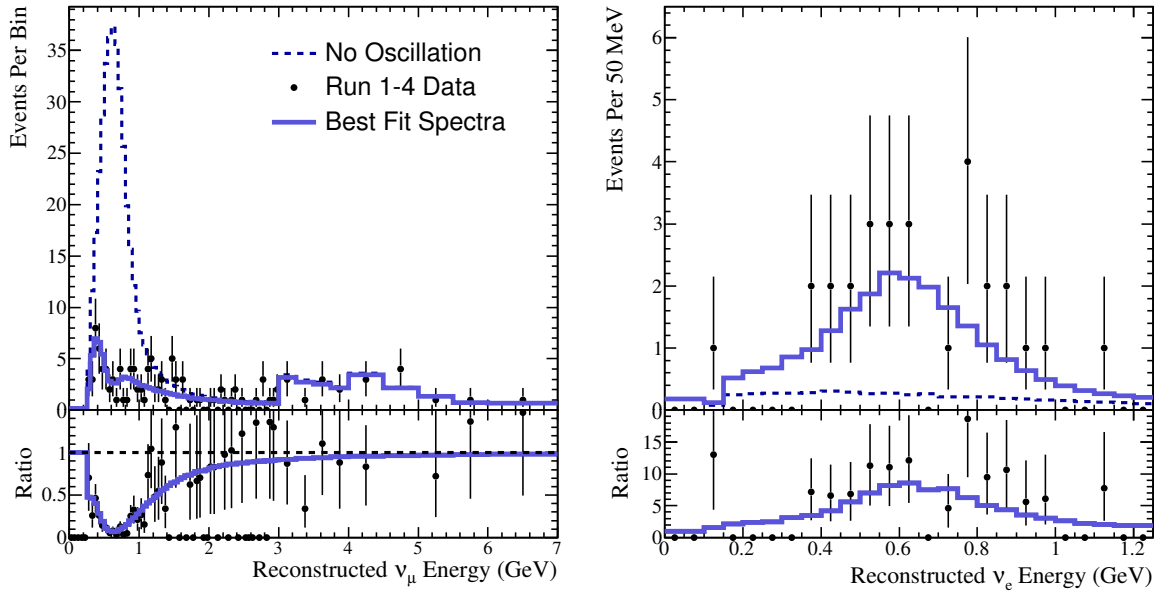


Figure 7.26.: Run 1–4 data best fit spectra for SuperK ν_μ and ν_e samples with reactor constraint.

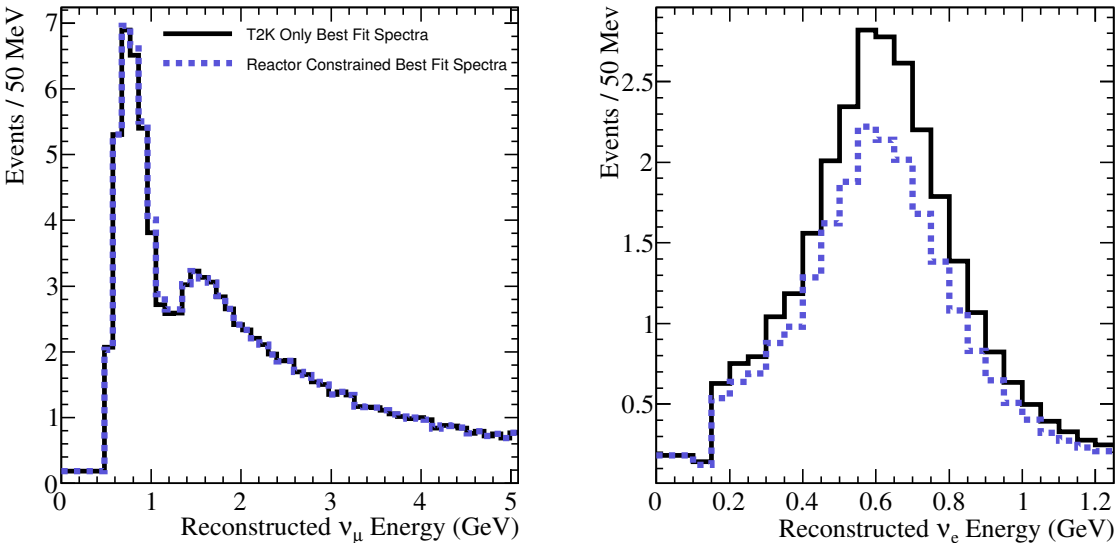


Figure 7.27.: Comparison of best fit spectra of T2K data with and without reactor constraint applied.

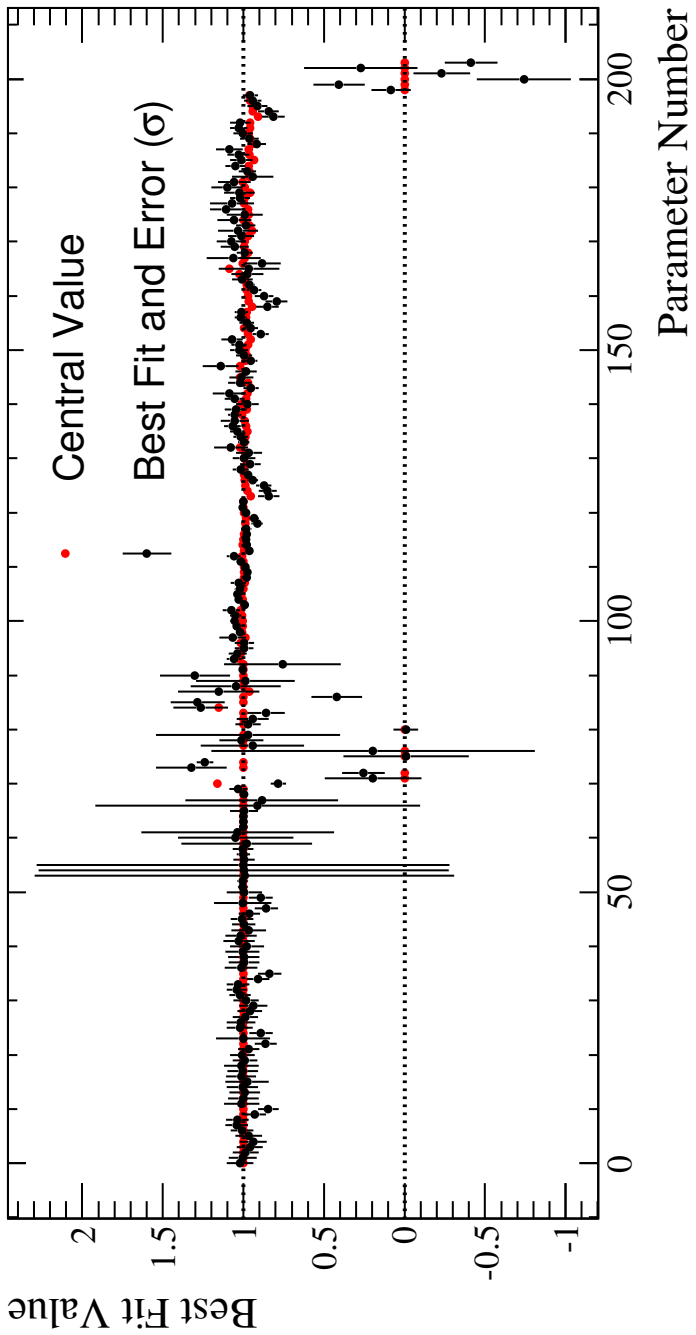


Figure 7.28.: Best fit values of the systematic parameters for the reactor constrained data fit. Shown also is the central value used in the constraint term. Parameter names can be found in Appendix A.

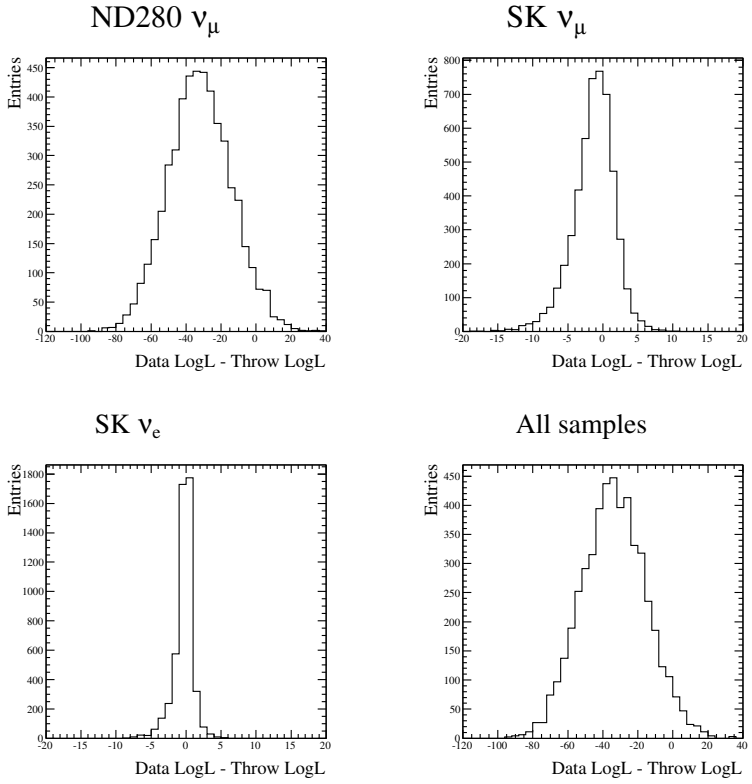


Figure 7.29.: Goodness-of-fit distributions for the three different samples in the fit and the summed total. The p-value is the percentage of each distribution which is greater than zero.

The goodness-of-fit was repeated for the reactor constrained data. Figure 7.29 shows the ND280, $1R_\mu$, $1R_e$, and total distributions for the quantity $\ln L_{data} - \ln L_{throw}$; the p-value is the percentage of this distribution above zero. The p-values are: ND280-only, 0.044; SK $1R_e$, 0.44; SK $1R_\mu$, 0.33; and all samples, 0.042. These values indicate no disagreement with data for the SK samples. It is interesting that the p-value for SK $1R_e$ increases slightly for this fit as compared to the T2K-only fit, despite the fact that the predicted number of events for the T2K-only fit is closer to the number of data events. This is due to the fact that the reactor constraint narrows the range of predicted distributions in the ensemble significantly, and therefore the predicted spectra from the throws do not move as far from the data as they do for the T2K-only fit.

The addition of the reactor constraint to the T2K data also produces some sensitivity in δ_{cp}. Figure 7.30 shows the δ_{cp} posterior for the mass prior P(NH)=1 (red line), the mass prior P(IH)=1 (blue line), and the mass prior P(NH)=P(IH)=0.5, with marginalization over the hierarchies (black line, with credible intervals show in greys). Figure 7.30

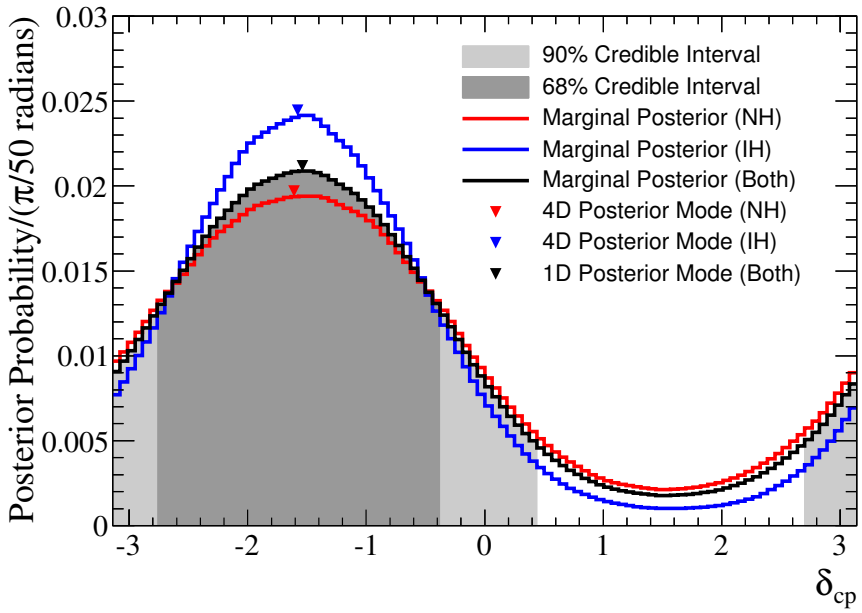


Figure 7.30.: The posterior probability for δ_{cp}, marginalized over all other parameters. The red curve shows the posterior for the mass prior P(NH)=1; the blue curve for the mass prior P(IH)=1; and the black curve for the mass prior P(NH)=P(IH)=0.5, marginalized over the hierarchies. The grey bands show the 68% and 90% credible intervals for the posterior corresponding to the black curve.

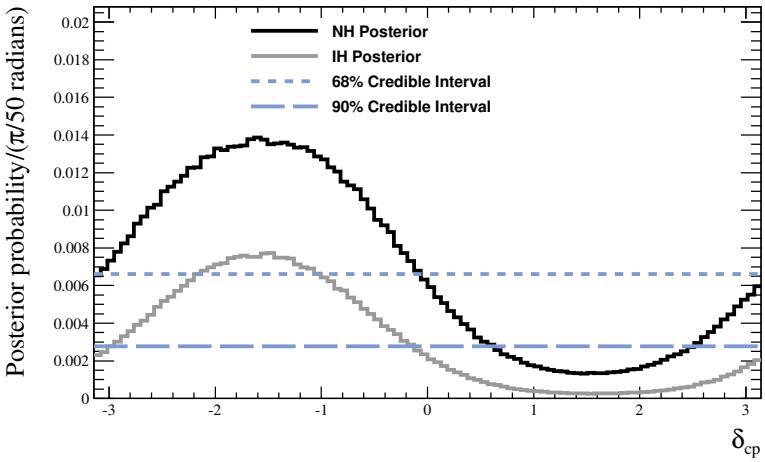


Figure 7.31.: The posterior probability for δ_{cp}, for the mass prior P(NH)=P(IH)=0.5, without marginalizing over the mass hierarchies. The dotted lines show the 68% and 90% credible intervals, where the allowed region is the region of the posterior above the line.

Table 7.6.: The 90% allowed credible interval for different methods of constructing the δ_{cp} posterior. The first line shows the interval boundaries in radians and the second in units of π.

Mass Prior and Method	90% Included Credible Interval
P(NH)=1	$[-\pi, 0.53] \cup [2.64, \pi]$
	$[-1, 0.17] \cup [0.84, 1]$
P(IH)=1	$[-\pi, 0.18] \cup [2.99, \pi]$
	$[-1, 0.056] \cup [0.95, 1]$
P(NH)=P(IH)=0.5, marg. hierarchy	$[-\pi, 0.43] \cup [2.74, \pi]$
	$[-1, 0.14] \cup [0.87, 1]$

shows that the choice of mass prior has only a small effect on the posterior distribution, indicating that the data has more constraining power, and that the experiment has a weak sensitivity to the mass hierarchy. Figure 7.31 shows the δ_{cp} posterior when considering the mass prior P(NH)=P(IH)=0.5 without marginalization over the hierarchies. Each of these methods answers a slightly different question about the preferred region for the value of δ_{cp}, and caution should be used when using these plots to describe them correctly. Table 7.6 enumerates the 90% allowed regions for the different methods.

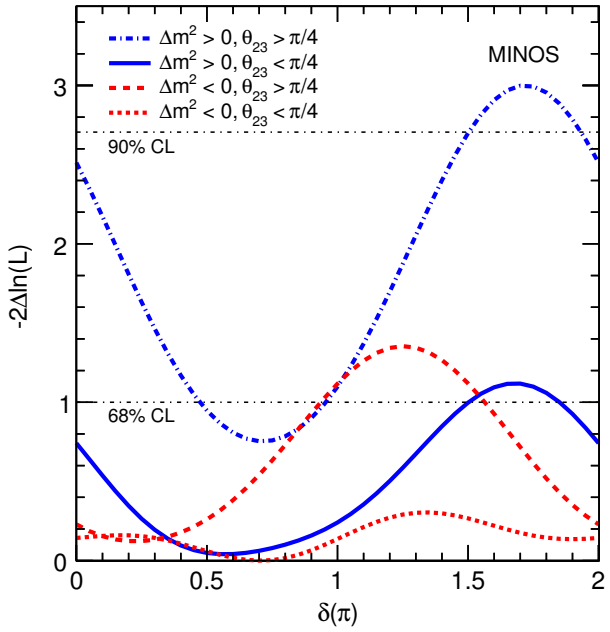


Figure 7.32.: MINOS likelihood curves for a data fit assuming different values of MH and sign of $\theta_{23} - \pi/4$. Figure taken from [109].

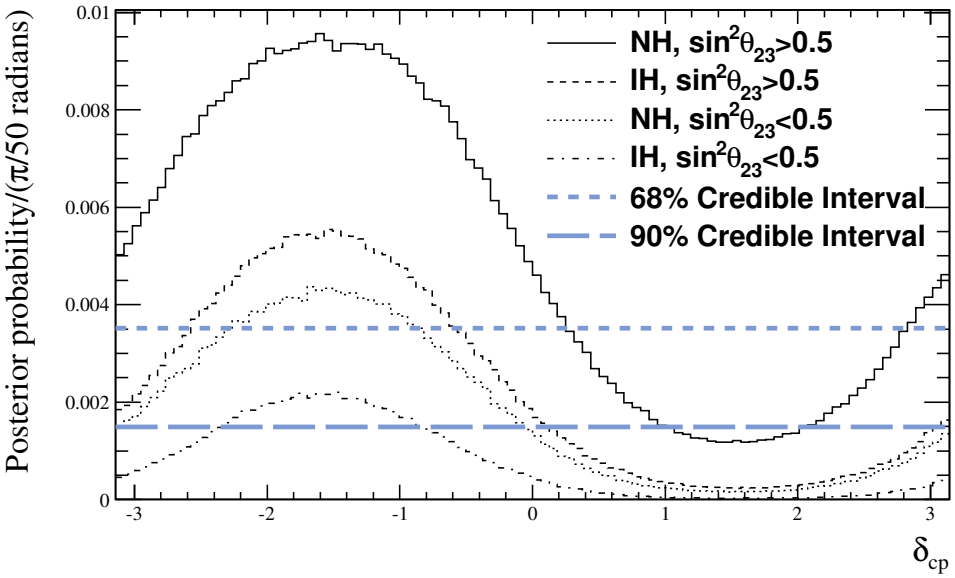


Figure 7.33.: The δ_{CP} posteriors with the priors P(NH)=P(IH)=0.5 and P(lower octant)=P(upper octant)=0.5, without marginalization over the hierarchy or the octant. The allowed region is the region of the posterior above the line.

Table 7.7.: Model comparison probabilities for normal and inverted mass hierarchies, as well as upper and lower octants.

	NH	IH	Sum
$\sin^2\theta_{23} \leq 0.5$	0.179	0.078	0.257
$\sin^2\theta_{23} > 0.5$	0.505	0.238	0.743
Sum	0.684	0.316	1.0

The constraint on δ_{CP} can also be considered separating the lower and upper octant as shown in Figure 7.33. Unlike the MINOS data (see Figure 7.32), the best fit point remains constant at $\approx -\pi/2$ for all of the choices of octant and hierarchy. However, some are more preferred than others; the inverted hierarchy/lower octant choice is excluded completely at the 68% level and nearly completely at the 90% level. By contrast, nearly all of the values of δ_{CP} in the normal hierarchy/upper octant choice are included in the 90% credible interval.

The Markov chain also provides an interesting and natural way to compare the mass hierarchies and octant probabilities. As discussed in Section 4.2.4, the comparison of two models can be made calculating the Bayes factor. Because all configurations of MH and octant are included in the posterior distribution, and the prior distributions for each model are the same, then the difficulty of calculating Equation 4.10 is significantly reduced as the first term is simply $p(M_i|I)/p(M_j|I) = 1$. This means that the Bayes factor, and the probability that the data support a particular combination of MH and octant is simply the proportion of steps of the chain in that region of phase space. Table 7.7 provides those probabilities for the two choices of mass hierarchy and two choices of octant. The interesting things to note are that the probability of the normal hierarchy (marginalized over octant) is 68.4% and the probability of the the upper octant (marginalized over hierarchy) is 74.3%. According to Table 4.2, the evidence of support for NH and upper octant is weak, and barely worth mentioning. Nevertheless, for the sake of completeness, it is given some attention.

7.3. Comparison to Other Experiments

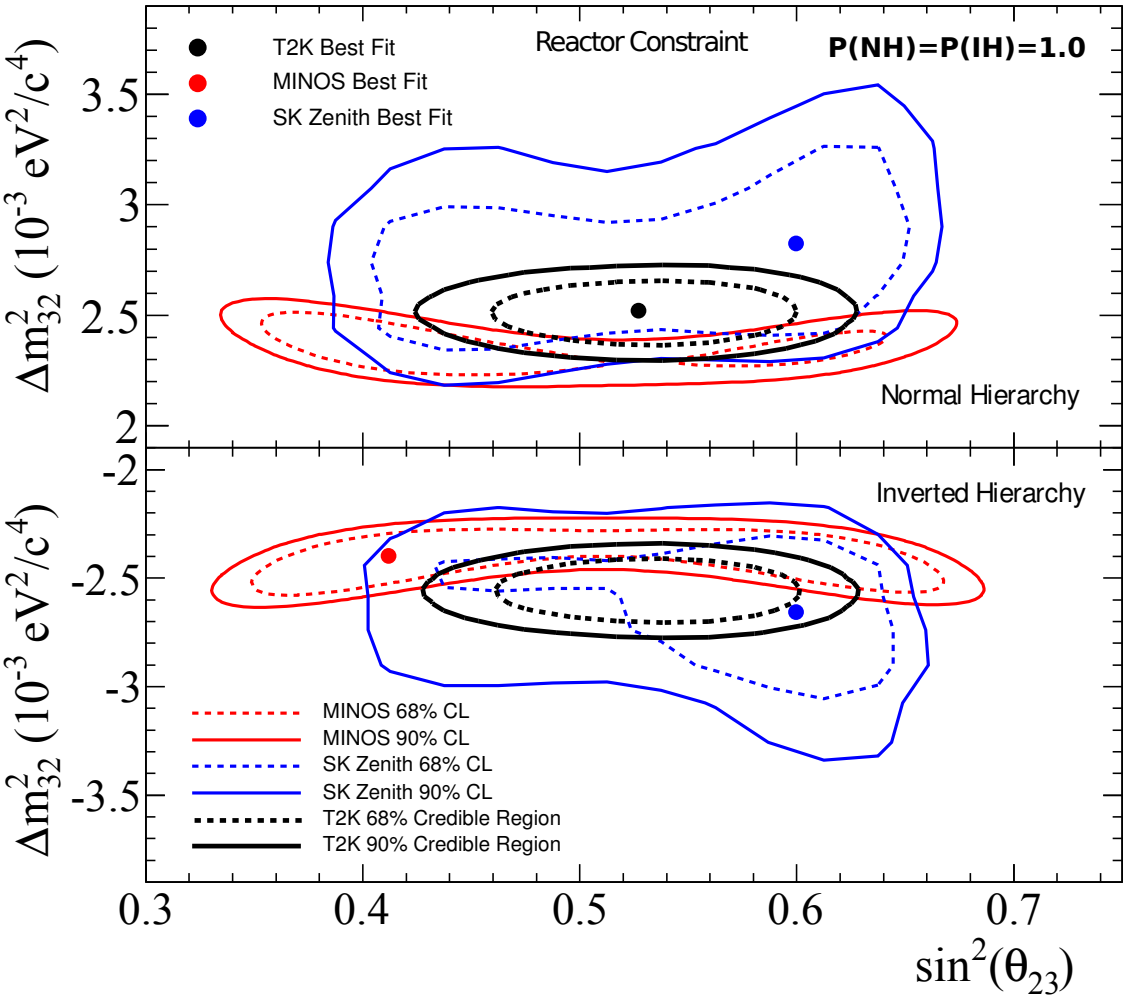


Figure 7.34.: Comparison of T2K reactor constrained result with SK Zenith and MINOS results.

Figure 7.34 shows a contour comparison of the reactor constrained analysis with recent SK [110] and MINOS [30] results. The confidence regions for SK were made using a fixed value of $\sin^2(2\theta_{13}) = 0.098$, and for MINOS using a fitted θ_{13} parameter constrained by $\sin^2\theta_{13} = 0.095 \pm 0.010$. The T2K credible intervals were made using the P(NH)=P(IH)=1.0 prior, in order to compare the contours more easily.

7.4. Posterior Parameter Correlation

Using the posterior distribution of the data analysis, it is possible to determine the covariance between all parameters of the analysis by constructing the 2-Dimensional marginal posterior of each parameter combination and calculating the covariance between the two axes. Figure 7.35 shows the posterior covariance matrix for all parameters, for the P(IH)=1 prior. The anti-correlation between the flux and ND280 detector systematics described in Figure 7.8 manifests itself in Figure 7.35.

The matrices shown in this section have had each element divided by the square root of their diagonal elements i.e. $\sqrt{V_{i,i}V_{j,j}}$, to enhance the visibility of the correlations between parameters. This limits the z axis between -1 and 1.

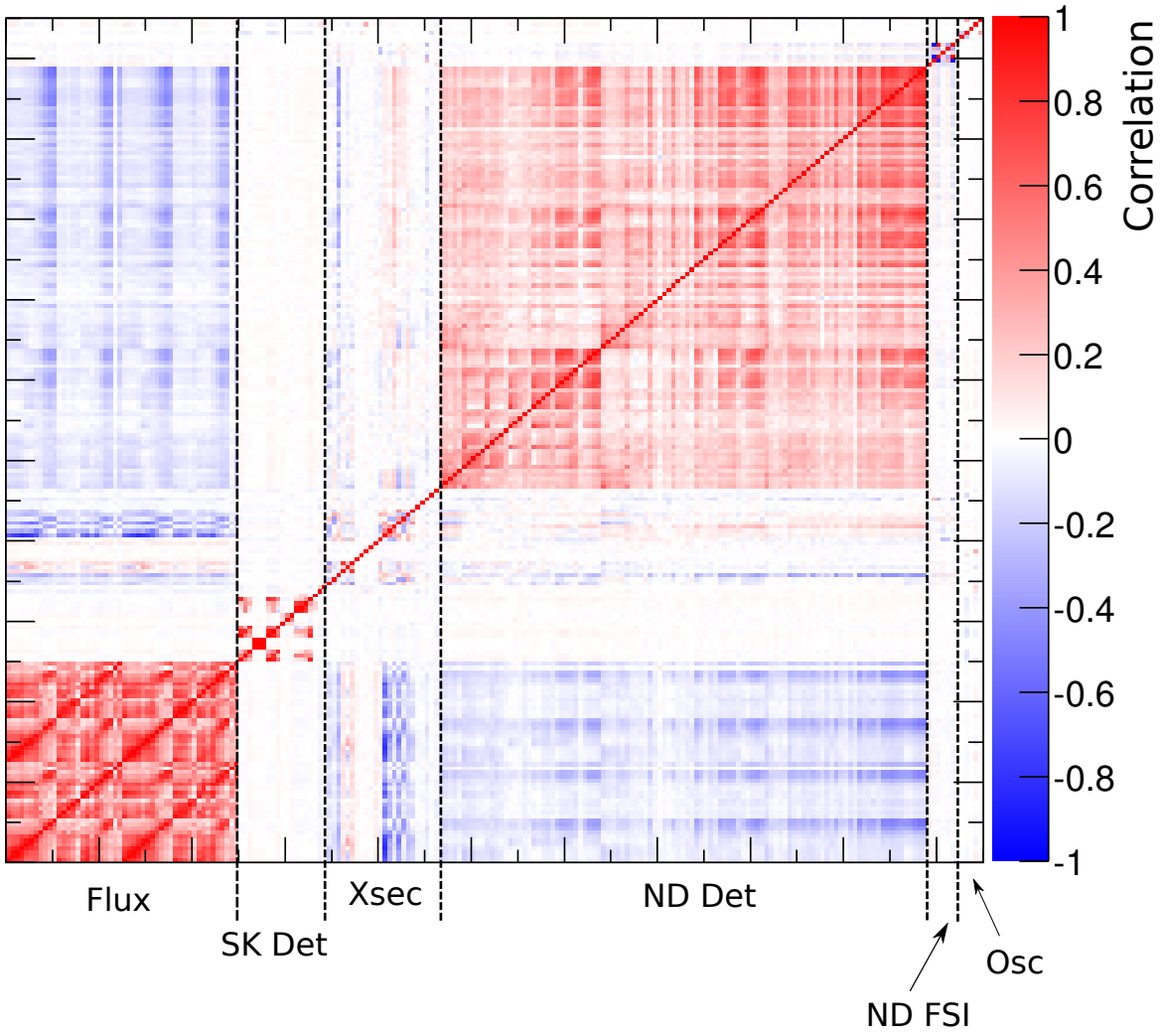


Figure 7.35.: Full posterior parameter correlation.

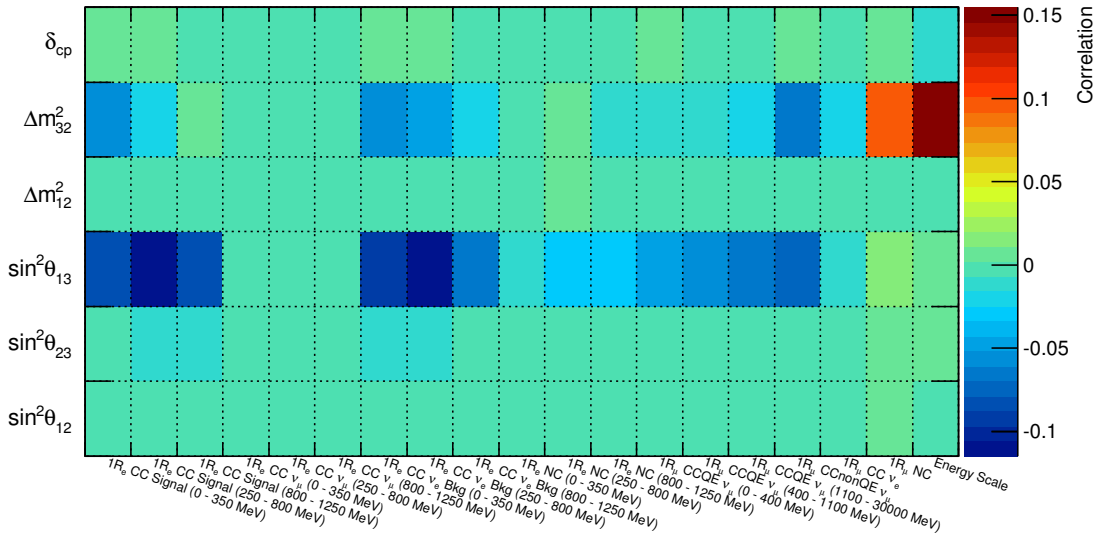


Figure 7.36.: Correlation between oscillation and SK detector parameters.

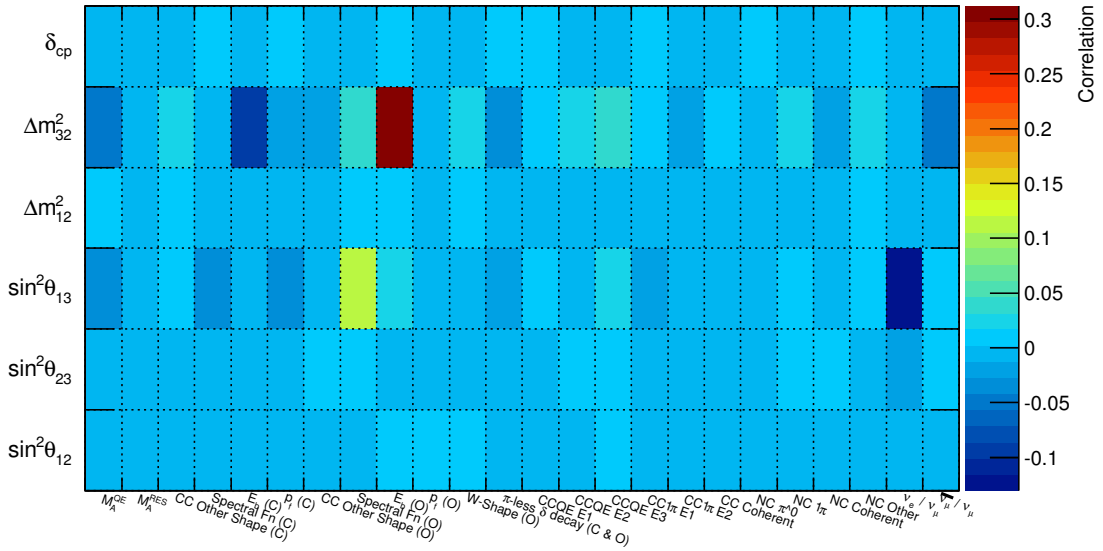


Figure 7.37.: Correlation between oscillation and cross section parameters.

Subsections of the full posterior covariance matrix are shown in Figures 7.36 and 7.37, which show the correlation between oscillation, SK detector and cross section parameters respectively. The correlations between the other sets of systematic parameters (flux, ND280 detector and FSI) and oscillation parameters are weak.

In general, the oscillation and flux parameters have no strong correlations. The SK energy scale and Δm^2_{32} parameters show correlation, which arises from the fact that these 2 parameters change the E_{rec} spectrum in similar ways. Further work to reduce the energy scale systematic would yield more sensitivity to Δm^2_{32}.

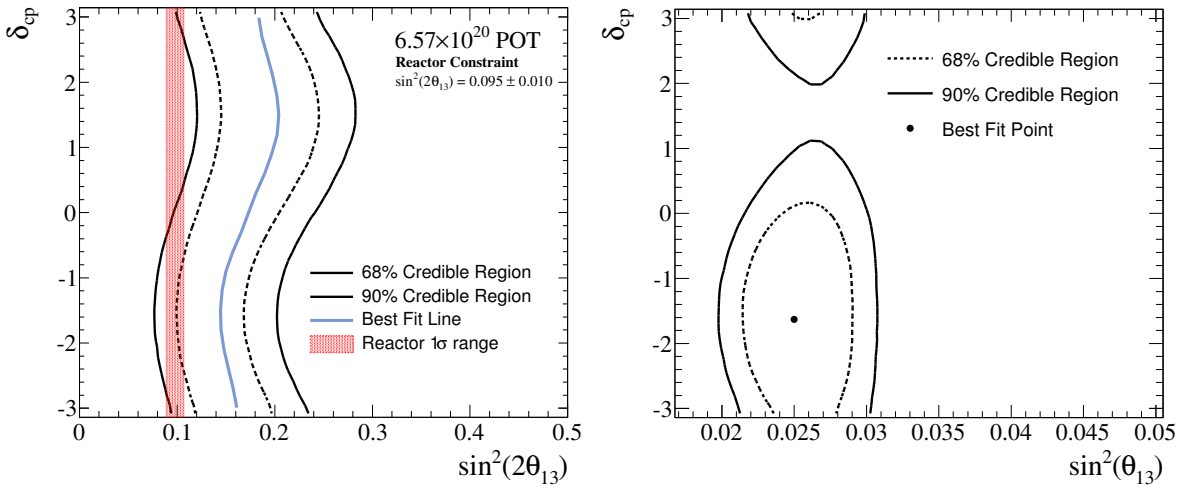
Likewise for the cross section parameters, there is a correlation between the binding energy on oxygen E_b and Δm^2_{32}. This motivates a measurement on the water component of the FGD at ND280 to help constrain Δm^2_{32}.

7.5. Summary

In summary, this chapter reports the results of a joint fit to the ND280 tracker ν_μ, SK 1R$_e$, and SK 1R$_\mu$. The best fit points for the T2K-only and reactor-constrained fits are shown in Table 7.8, where the errors given on the individual parameters come from the 1D 68% credible intervals. The 68% and 90% contours for the δ_{CP}-sin$^2\,\theta_{13}$ and Δm^2_{32}-sin$^2\,\theta_{23}$ parameter spaces are shown in Figures 7.38 and 7.39, respectively. With the reactor constraint, the 90% included credible interval is $[-1, 0.12]$ and $[0.89, 1]$, in units of π, which is shown in Figure 7.40.

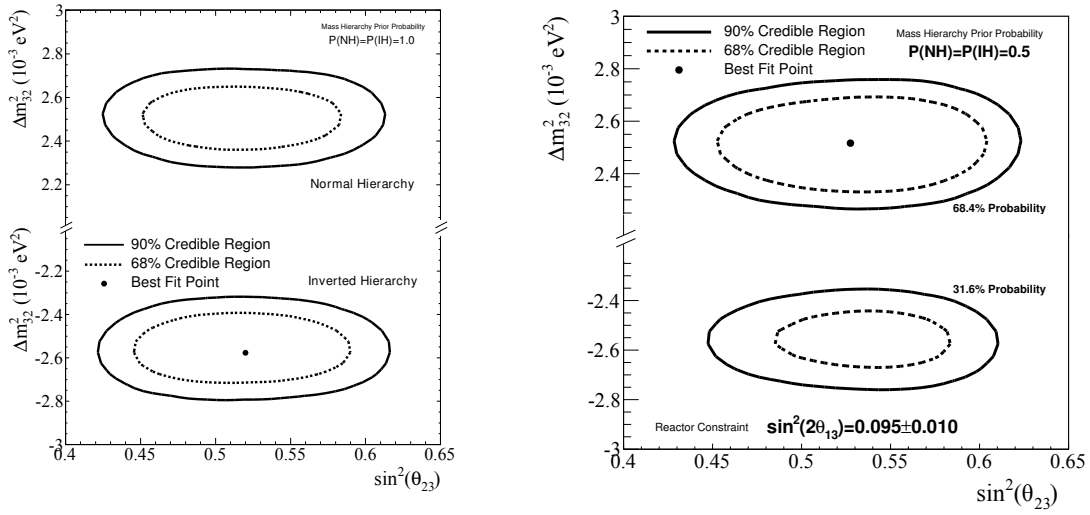
Table 7.8.: Best-fit values for oscillation parameters extracted from the marginal posterior of the Run 1–4 data. Negative sign on Δm^2_{32} indicates inverted hierarchy is preferred, positive indicates normal hierarchy. No error is given for δ_{cp} for the reactor-constrained fit, as the distribution is decidedly non-gaussian, and would be misleading.

	Δm^2_{32}	$\sin^2(\theta_{23})$	$\sin^2(\theta_{13})$	δ_{cp}
T2K-only	$-2.57 \pm 0.11 \times 10^{-3}$ eV^2	$0.520^{+0.045}_{-0.050}$	$0.0454^{+0.011}_{-0.014}$	0 (fixed)
Reactor-constrained	$2.51 \pm 0.11 \times 10^{-3}$ eV^2	$0.528^{+0.055}_{-0.038}$	0.0250 ± 0.0026	-1.601



(a) (b)

Figure 7.38.: Run 1–4 data fit 2D contours in δ_{CP}–sin$^2(\theta_{13})$ space, using the mass prior P(NH)=P(IH)=0.5. The plots in (a) are for the T2K-only fit and in (b) for the reactor-constrained fit.



(a) **(b)**

Figure 7.39.: Run 1–4 data fit 2D contours in Δm^2_{23}–$\sin^2(\theta_{23})$ space, using the mass prior P(NH)=P(IH)=0.5. The plots in (a) are for the T2K-only fit and in (b) for the reactor-constrained fit.

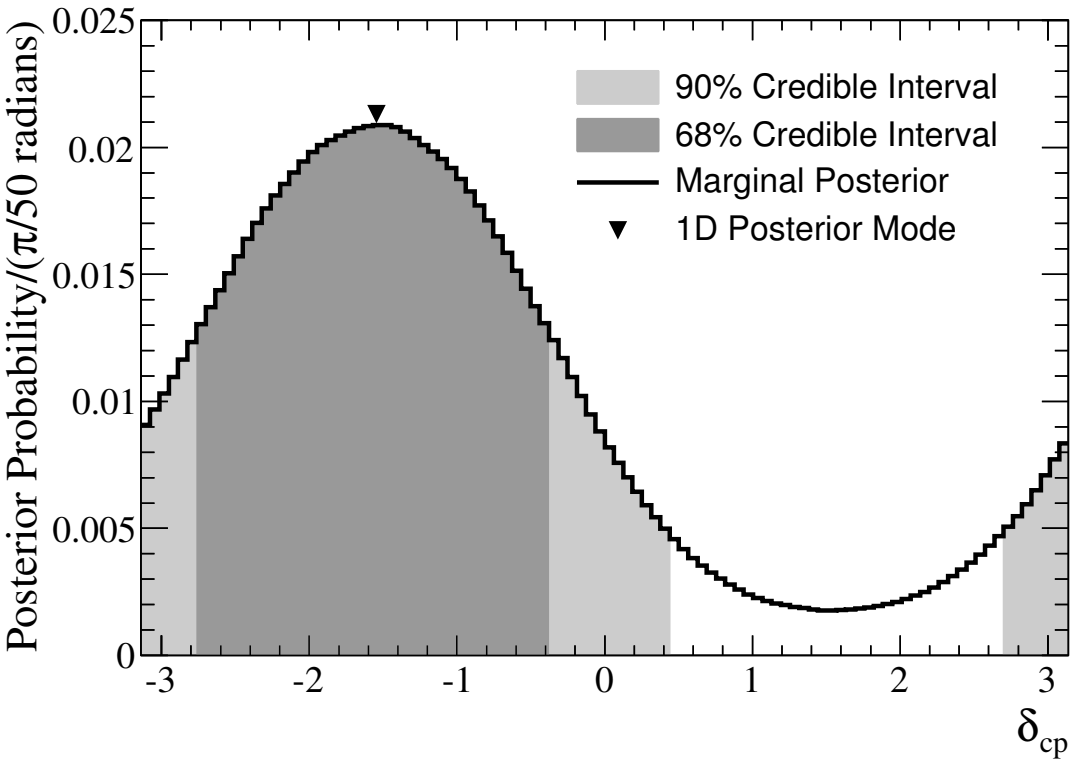


Figure 7.40.: Run 1–4 data fit 1D credible interval for δ_{CP}, using the mass prior P(NH)=P(IH)=0.5 and the reactor constraint.

Chapter 8.

Conclusions and Discussions

A Bayesian statistical analysis of Run 1–4 data from T2K near and far detectors was performed. The simultaneous analysis of near detector data provides constraint to the dominant systematics in the oscillation measurement, whilst the joint analysis of both ν_μ and ν_e selections at the far detector give sensitivity to the atmospheric mixing parameters through the disappearance channel, and sensitivity to θ_{13} and δ_{CP} through the appearance channel, especially when combined with the reactor constraint prior.

The analysis made use of a Markov chain Monte Carlo to perform the full marginalization of 206 nuisance parameters. All PDFs were constructed using an event-by-event reweighting method, which made use of GPU technology which was essential to the feasibility of the technique.

The results of the analysis with T2K Run 1–4 data, reported at $\delta_{CP} = 0$, are

$$\Delta m^2_{32} = -2.57 \pm 0.11 \times 10^{-3}\ \text{eV}^2,\ \sin^2\theta_{23} = 0.520^{+0.045}_{-0.050},\ \text{and}\ \sin^2\theta_{13} = 0.0454^{+0.011}_{-0.014}.$$

With the inclusion of the reactor prior constraint for the value of θ_{13}, the results are

$$\Delta m^2_{32} = 2.51 \pm 0.11 \times 10^{-3}\ \text{eV}^2,\ \sin^2\theta_{23} = 0.528^{+0.055}_{-0.038},\ \text{and}\ \delta_{CP} = -1.587.$$

As the reactor constraint introduces sensitivity to the δ_{CP} parameter, the reactor constrained analysis finds that the 90% credible interval for δ_{CP} excludes 0.14–0.87, in units of π. The combination of T2K and reactor experimental data prefers the normal mass hierarchy, however this preference is not strong and does not warrant much attention.

These results are consistent with other T2K oscillation analyses, and also competitive with other experimental results. Sensitivity is expected to further increase as T2K collects

more POT, and as the reactor experiments further constrain the value of θ_{13}, the apparent tension between measured values of θ_{13} by both experiments can give useful insight into the value of δ_{CP}; if the trend continues then it is possible that together T2K and reactor experiments could exclude non-zero δ_{CP}. T2K plans to operate an anti-neutrino beam, which opens the possibility for looking at CP-violation in the lepton sector with only T2K data.

The simultaneous analysis of both near and far detector data has merit in that it has proven to be a validation of the previous method of extrapolating the near detector constraint as a multivariate Gaussian (at least, with current statistics), but also has acted as a proof-of-concept for furthering the capabilities of ND280. As discussed in Section 7.4, ND280 has the potential to constrain SK detector systematics, and with the simultaneous method the assumption of Gaussian uncertainties need not be used.

The simultaneous analysis of five samples from two detectors results in over 200 parameters. A frequentist approach of minimizing and profiling these parameters according to the likelihood function does not fare well with such large numbers of parameters, and presents a limit on the scope of a single analysis. Should additional samples be added to the oscillation analyses, the MCMC method is a natural choice as the number of dimensions scales linearly with the time taken to perform the analysis. The MCMC method also intrinsically marginalizes all nuisance parameters.

This analysis presents all results in the form of Bayesian credible intervals, which were defined in Section 4.2.3. It is apparent that the HEP community is more familiar with the frequentist alternative called confidence intervals, and often times they are interpreted as one another and even compared directly (rather shamefully, this occurs within this very thesis!). This is not recommended as the two methods present different definitions of intervals; a *confidence interval* represents the region that would contain say 90% of best-fit points should the experiment be repeated 100 times. In contrast, a *credible interval* represents the actual probability distribution of the value of a parameter given the single data set. The 90% credible region is telling us that the data is 90% sure the true value of the parameter lies within its bounds, and thus there is no question of adequate coverage associated with frequentist contours.

The analysis has several improvements that could be made, namely by changing from using reconstructed energy to using momentum-angle kinematic variables for SK PDFs, which should retain more information, as is used for the ν_e appearance analysis [76]. Although this analysis uses an event-by-event reweighting method, many of the systematic

uncertainties still behave as though they were binned. The response functions for SK cross-section uncertainties could be moved to an event-by-event method due to the GPU acceleration technique presented here. This would model the non-linear cross-section model with more precision, which was prohibited in the past due to speed concerns.

As SK has been running for much longer than the T2K experiment, it has a wealth of atmospheric neutrino data. The simultaneous analysis of atmospheric neutrino and beam neutrino data would yield not only further constraint on the atmospheric mixing parameters, but also perhaps hint at neutrino mass hierarchy, due to the long baselines associated with atmospheric neutrinos. SK and T2K are prime candidates for a combined atmospheric and beam analysis due to sharing the same detector and thus have highly correlated detector systematics.

Appendix A.

Parameter Index

Parameter enumeration used in Chapter 7 can be found in this Appendix. The first number lists the absolute parameter number, and the second bold-face number refers to the parameter number relative to the parameter group to which it belongs. There are a total of 210 parameters used in the analysis.

Flux Errors

0. **0** ND280 ν_μ (0 - 400 MeV)
1. **1** ND280 ν_μ (400 - 500 MeV)
2. **2** ND280 ν_μ (500 - 600 MeV)
3. **3** ND280 ν_μ (600 - 700 MeV)
4. **4** ND280 ν_μ (700 - 1000 MeV)
5. **5** ND280 ν_μ (1000 - 1500 MeV)
6. **6** ND280 ν_μ (1500 - 2500 MeV)
7. **7** ND280 ν_μ (2500 - 3500 MeV)
8. **8** ND280 ν_μ (3500 - 5000 MeV)
9. **9** ND280 ν_μ (5000 - 7000 MeV)
10. **10** ND280 ν_μ (7000 - 30000 MeV)
11. **11** ND280 anti-ν_μ (0 - 700 MeV)
12. **12** ND280 anti-ν_μ (700 - 1000 MeV)
13. **13** ND280 anti-ν_μ (1000 - 1500 MeV)
14. **14** ND280 anti-ν_μ (1500 - 2500 MeV)
15. **15** ND280 anti-ν_μ (2500 - 30000 MeV)
16. **16** ND280 ν_e (0 - 500 MeV)
17. **17** ND280 ν_e (500 - 700 MeV)
18. **18** ND280 ν_e (700 - 800 MeV)
19. **19** ND280 ν_e (800 - 1500 MeV)
20. **20** ND280 ν_e (1500 - 2500 MeV)
21. **21** ND280 ν_e (2500 - 4000 MeV)
22. **22** ND280 ν_e (4000 - 30000 MeV)
23. **23** ND280 anti-ν_e (0 - 2500 MeV)
24. **24** ND280 anti-ν_e (2500 - 30000 MeV)
25. **25** SK ν_μ (0 - 400 MeV)
26. **26** SK ν_μ (400 - 500 MeV)
27. **27** SK ν_μ (500 - 600 MeV)
28. **28** SK ν_μ (600 - 700 MeV)
29. **29** SK ν_μ (700 - 1000 MeV)
30. **30** SK ν_μ (1000 - 1500 MeV)
31. **31** SK ν_μ (1500 - 2500 MeV)
32. **32** SK ν_μ (2500 - 3500 MeV)
33. **33** SK ν_μ (3500 - 5000 MeV)
34. **34** SK ν_μ (5000 - 7000 MeV)
35. **35** SK ν_μ (7000 - 30000 MeV)
36. **36** SK anti-ν_μ (0 - 700 MeV)
37. **37** SK anti-ν_μ (700 - 1000 MeV)
38. **38** SK anti-ν_μ (1000 - 1500 MeV)
39. **39** SK anti-ν_μ (1500 - 2500 MeV)
40. **40** SK anti-ν_μ (2500 - 30000 MeV)
41. **41** SK ν_e (0 - 500 MeV)
42. **42** SK ν_e (500 - 700 MeV)
43. **43** SK ν_e (700 - 800 MeV)
44. **44** SK ν_e (800 - 1500 MeV)
45. **45** SK ν_e (1500 - 2500 MeV)
46. **46** SK ν_e (2500 - 4000 MeV)
47. **47** SK ν_e (4000 - 30000 MeV)
48. **48** SK anti-ν_e (0 - 2500 MeV)
49. **49** SK anti-ν_e (2500 - 30000 MeV)

SK Detector Errors

50. **0** 1R$_e$ CC Signal (0 - 350 MeV)
51. **1** 1R$_e$ CC Signal (350 - 800 MeV)
52. **2** 1R$_e$ CC Signal (800 - 1250 MeV)
53. **3** 1R$_e$ CC ν_μ (0 - 350 MeV)
54. **4** 1R$_e$ CC ν_μ (350 - 800 MeV)
55. **5** 1R$_e$ CC ν_μ (800 - 1250 MeV)
56. **6** 1R$_e$ CC ν_e Bkg (0 - 350 MeV)
57. **7** 1R$_e$ CC ν_e Bkg (350 - 800 MeV)
58. **8** 1R$_e$ CC ν_e Bkg (800 - 1250 MeV)
59. **9** 1R$_e$ NC (0 - 350 MeV)
60. **10** 1R$_e$ NC (350 - 800 MeV)
61. **11** 1R$_e$ NC (800 - 1250 MeV)
62. **12** 1R$_\mu$ CCQE ν_μ (0 - 400 MeV)
63. **13** 1R$_\mu$ CCQE ν_μ (400 - 1100 MeV)
64. **14** 1R$_\mu$ CCQE ν_μ (1100 - 30000 MeV)
65. **15** 1R$_\mu$ CCnonQE ν_μ
66. **16** 1R$_\mu$ CC ν_e
67. **17** 1R$_\mu$ NC
68. **18** Energy Scale

Cross-Section Errors

69. **0** M_A^{QE}
70. **1** M_A^{RES}
71. **2** CC Other Shape (C)
72. **3** Spectral Function (C)
73. **4** Binding Energy (C)
74. **5** Fermi Momentum (C)
75. **6** CC Other Shape (O)
76. **7** Spectral Function (O)

77. **8** Binding Energy (O)
78. **9** Fermi Momentum (O)
79. **10** W-Shape (O)
80. **11** Pion-less Delta Decay (C & O)
81. **12** CCQE E1 (0 - 1500 MeV)
82. **13** CCQE E2 (1500 - 3500 MeV)
83. **14** CCQE E3 (> 3500 MeV)
84. **15** CC1pi E1 (0 - 2500 MeV)
85. **16** CC1pi E2 (> 2500 MeV)
86. **17** CC Coherent
87. **18** NCπ^0
88. **19** NC1π
89. **20** NC Coherent
90. **21** NC Other
91. **22** ν_e / ν_μ
92. **23** anti-ν_μ / ν_μ

ND280 Detector Errors

93. **0** CC0pi - p = 0 - 300 MeV, $\cos\theta$ = -1.0 - 0.85
94. **1** CC0pi - p = 0 - 300 MeV, $\cos\theta$ = 0.85 - 0.9
95. **2** CC0pi - p = 0 - 300 MeV, $\cos\theta$ = 0.9 - 0.94
96. **3** CC0pi - p = 0 - 300 MeV, $\cos\theta$ = 0.94 - 0.98
97. **4** CC0pi - p = 0 - 300 MeV, $\cos\theta$ = 0.98 - 1.0
98. **5** CC0pi - p = 300 - 500 MeV, $\cos\theta$ = -1.0 - 0.85
99. **6** CC0pi - p = 300 - 500 MeV, $\cos\theta$ = 0.85 - 0.9
100. **7** CC0pi - p = 300 - 500 MeV, $\cos\theta$ = 0.9 - 0.94
101. **8** CC0pi - p = 300 - 500 MeV, $\cos\theta$ = 0.94 - 0.98
102. **9** CC0pi - p = 300 - 500 MeV, $\cos\theta$ = 0.98 - 1.0
103. **10** CC0pi - p = 500 - 600 MeV, $\cos\theta$ = -1.0 - 0.85
104. **11** CC0pi - p = 500 - 600 MeV, $\cos\theta$ = 0.85 - 0.9
105. **12** CC0pi - p = 500 - 600 MeV, $\cos\theta$ = 0.9 - 0.94
106. **13** CC0pi - p = 500 - 600 MeV, $\cos\theta$ = 0.94 - 0.98
107. **14** CC0pi - p = 500 - 600 MeV, $\cos\theta$ = 0.98 - 1.0
108. **15** CC0pi - p = 600 - 700 MeV, $\cos\theta$ = -1.0 - 0.85
109. **16** CC0pi - p = 600 - 700 MeV, $\cos\theta$ = 0.85 - 0.9
110. **17** CC0pi - p = 600 - 700 MeV, $\cos\theta$ = 0.9 - 0.94
111. **18** CC0pi - p = 600 - 700 MeV, $\cos\theta$ = 0.94 - 0.98
112. **19** CC0pi - p = 600 - 700 MeV, $\cos\theta$ = 0.98 - 1.0
113. **20** CC0pi - p = 700 - 1000 MeV, $\cos\theta$ = -1.0 - 0.85
114. **21** CC0pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.85 - 0.9
115. **22** CC0pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.9 - 0.94
116. **23** CC0pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.94 - 0.98
117. **24** CC0pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.98 - 1.0
118. **25** CC0pi - p = 1000 - 2000 MeV, $\cos\theta$ = -1.0 - 0.85
119. **26** CC0pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.85 - 0.9
120. **27** CC0pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.9 - 0.94
121. **28** CC0pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.94 - 0.98
122. **29** CC0pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.98 - 1.0
123. **30** CC0pi - p = 2000 - 30000 MeV, $\cos\theta$ = -1.0 - 0.85
124. **31** CC0pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.85 - 0.9
125. **32** CC0pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.9 - 0.94
126. **33** CC0pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.94 - 0.98
127. **34** CC0pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.98 - 1.0
128. **35** CC1pi - p = 0 - 300 MeV, $\cos\theta$ = -1.0 - 0.85
129. **36** CC1pi - p = 0 - 300 MeV, $\cos\theta$ = 0.85 - 0.9
130. **37** CC1pi - p = 0 - 300 MeV, $\cos\theta$ = 0.9 - 0.94
131. **38** CC1pi - p = 0 - 300 MeV, $\cos\theta$ = 0.94 - 0.98
132. **39** CC1pi - p = 0 - 300 MeV, $\cos\theta$ = 0.98 - 1.0
133. **40** CC1pi - p = 300 - 500 MeV, $\cos\theta$ = -1.0 - 0.85
134. **41** CC1pi - p = 300 - 500 MeV, $\cos\theta$ = 0.85 - 0.9
135. **42** CC1pi - p = 300 - 500 MeV, $\cos\theta$ = 0.9 - 0.94
136. **43** CC1pi - p = 300 - 500 MeV, $\cos\theta$ = 0.94 - 0.98
137. **44** CC1pi - p = 300 - 500 MeV, $\cos\theta$ = 0.98 - 1.0
138. **45** CC1pi - p = 500 - 600 MeV, $\cos\theta$ = -1.0 - 0.85
139. **46** CC1pi - p = 500 - 600 MeV, $\cos\theta$ = 0.85 - 0.9
140. **47** CC1pi - p = 500 - 600 MeV, $\cos\theta$ = 0.9 - 0.94
141. **48** CC1pi - p = 500 - 600 MeV, $\cos\theta$ = 0.94 - 0.98
142. **49** CC1pi - p = 500 - 600 MeV, $\cos\theta$ = 0.98 - 1.0
143. **50** CC1pi - p = 600 - 700 MeV, $\cos\theta$ = -1.0 - 0.85
144. **51** CC1pi - p = 600 - 700 MeV, $\cos\theta$ = 0.85 - 0.9
145. **52** CC1pi - p = 600 - 700 MeV, $\cos\theta$ = 0.9 - 0.94
146. **53** CC1pi - p = 600 - 700 MeV, $\cos\theta$ = 0.94 - 0.98
147. **54** CC1pi - p = 600 - 700 MeV, $\cos\theta$ = 0.98 - 1.0
148. **55** CC1pi - p = 700 - 1000 MeV, $\cos\theta$ = -1.0 - 0.85
149. **56** CC1pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.85 - 0.9
150. **57** CC1pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.9 - 0.94
151. **58** CC1pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.94 - 0.98
152. **59** CC1pi - p = 700 - 1000 MeV, $\cos\theta$ = 0.98 - 1.0
153. **60** CC1pi - p = 1000 - 2000 MeV, $\cos\theta$ = -1.0 - 0.85
154. **61** CC1pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.85 - 0.9
155. **62** CC1pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.9 - 0.94
156. **63** CC1pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.94 - 0.98
157. **64** CC1pi - p = 1000 - 2000 MeV, $\cos\theta$ = 0.98 - 1.0
158. **65** CC1pi - p = 2000 - 30000 MeV, $\cos\theta$ = -1.0 - 0.85
159. **66** CC1pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.85 - 0.9
160. **67** CC1pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.9 - 0.94
161. **68** CC1pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.94 - 0.98
162. **69** CC1pi - p = 2000 - 30000 MeV, $\cos\theta$ = 0.98 - 1.0
163. **70** CCother - p = 0 - 300 MeV, $\cos\theta$ = -1.0 - 0.85
164. **71** CCother - p = 0 - 300 MeV, $\cos\theta$ = 0.85 - 0.9
165. **72** CCother - p = 0 - 300 MeV, $\cos\theta$ = 0.9 - 0.94
166. **73** CCother - p = 0 - 300 MeV, $\cos\theta$ = 0.94 - 0.98
167. **74** CCother - p = 0 - 300 MeV, $\cos\theta$ = 0.98 - 1.0
168. **75** CCother - p = 300 - 500 MeV, $\cos\theta$ = -1.0 - 0.85
169. **76** CCother - p = 300 - 500 MeV, $\cos\theta$ = 0.85 - 0.9
170. **77** CCother - p = 300 - 500 MeV, $\cos\theta$ = 0.9 - 0.94
171. **78** CCother - p = 300 - 500 MeV, $\cos\theta$ = 0.94 - 0.98
172. **79** CCother - p = 300 - 500 MeV, $\cos\theta$ = 0.98 - 1.0
173. **80** CCother - p = 500 - 600 MeV, $\cos\theta$ = -1.0 - 0.85
174. **81** CCother - p = 500 - 600 MeV, $\cos\theta$ = 0.85 - 0.9
175. **82** CCother - p = 500 - 600 MeV, $\cos\theta$ = 0.9 - 0.94
176. **83** CCother - p = 500 - 600 MeV, $\cos\theta$ = 0.94 - 0.98
177. **84** CCother - p = 500 - 600 MeV, $\cos\theta$ = 0.98 - 1.0
178. **85** CCother - p = 600 - 700 MeV, $\cos\theta$ = -1.0 - 0.85
179. **86** CCother - p = 600 - 700 MeV, $\cos\theta$ = 0.85 - 0.9
180. **87** CCother - p = 600 - 700 MeV, $\cos\theta$ = 0.9 - 0.94
181. **88** CCother - p = 600 - 700 MeV, $\cos\theta$ = 0.94 - 0.98
182. **89** CCother - p = 600 - 700 MeV, $\cos\theta$ = 0.98 - 1.0
183. **90** CCother - p = 700 - 1000 MeV, $\cos\theta$ = -1.0 - 0.85
184. **91** CCother - p = 700 - 1000 MeV, $\cos\theta$ = 0.85 - 0.9
185. **92** CCother - p = 700 - 1000 MeV, $\cos\theta$ = 0.9 - 0.94
186. **93** CCother - p = 700 - 1000 MeV, $\cos\theta$ = 0.94 - 0.98
187. **94** CCother - p = 700 - 1000 MeV, $\cos\theta$ = 0.98 - 1.0
188. **95** CCother - p = 1000 - 2000 MeV, $\cos\theta$ = -1.0 - 0.85
189. **96** CCother - p = 1000 - 2000 MeV, $\cos\theta$ = 0.85 - 0.9
190. **97** CCother - p = 1000 - 2000 MeV, $\cos\theta$ = 0.9 - 0.94
191. **98** CCother - p = 1000 - 2000 MeV, $\cos\theta$ = 0.94 - 0.98
192. **99** CCother - p = 1000 - 2000 MeV, $\cos\theta$ = 0.98 - 1.0
193. **100** CCother - p = 2000 - 30000 MeV, $\cos\theta$ = -1.0 - 0.85
194. **101** CCother - p = 2000 - 30000 MeV, $\cos\theta$ = 0.85 - 0.9
195. **102** CCother - p = 2000 - 30000 MeV, $\cos\theta$ = 0.9 - 0.94
196. **103** CCother - p = 2000 - 30000 MeV, $\cos\theta$ = 0.94 - 0.98
197. **104** CCother - p = 2000 - 30000 MeV, $\cos\theta$ = 0.98 - 1.0

ND280 FSI Errors

198. **0** INEL LO (ND280 FSI)
199. **1** INEL HI (ND280 FSI)
200. **2** PION PROD (ND280 FSI)
201. **3** PION ABS (ND280 FSI)
202. **4** CEX LO (ND280 FSI)
203. **5** CEX HI (ND280 FSI)

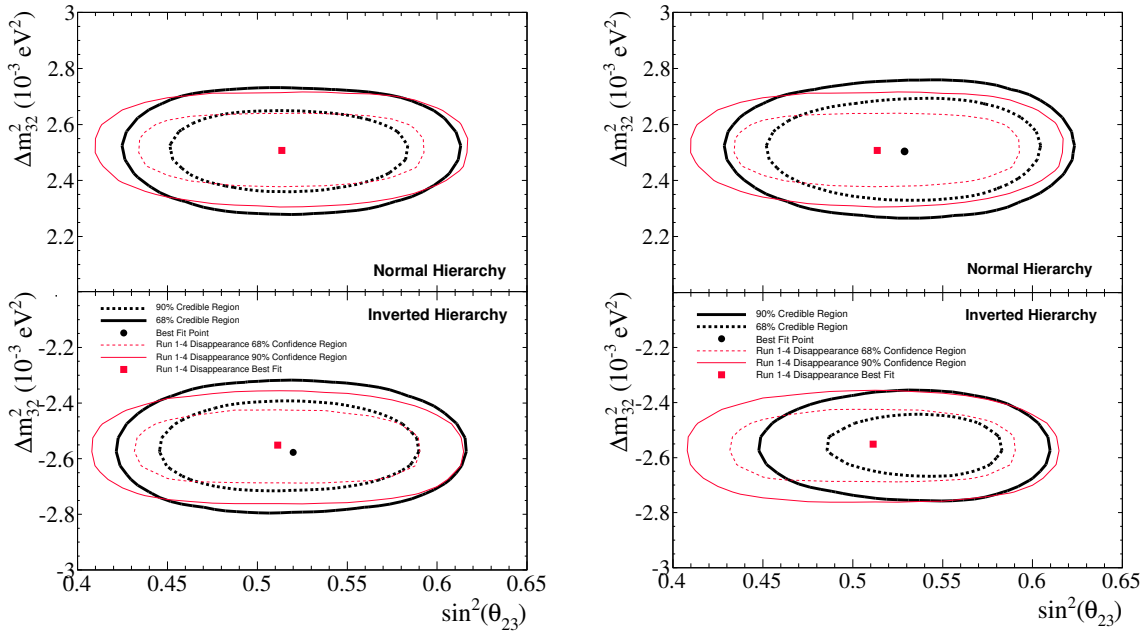
Oscillation Parameters

204. **0** $\sin^2\theta_{12}$
205. **1** $\sin^2\theta_{23}$
206. **2** $\sin^2\theta_{13}$
207. **3** Δm^2_{12}
208. **4** Δm^2_{23}
209. **5** δ_{cp}

Appendix B.

Comparison to Run 1–4 Disappearance Results

Figure B.1 shows the results of this fit in the Δm^2_{32}–$\sin^2\theta_{23}$ space compared to the Feldman-Cousins contours for the Run 1–4 disappearance analysis published by T2K [32]. Although it is not statistically valid to compare credible regions with confidence regions, a qualitative summary can be achieved, in that the inclusion of the reactor constraint for θ_{13} restricts the θ_{23} phase space such that more sensitivity is gained.



(a) **(b)**

Figure B.1.: Comparison of the results of this analysis to the 2013 Run 1–4 disappearance only analysis. Results are shown using the P(NH)=P(IH)=0.5 mass prior. T2K-only results are shown in (a) and reactor-constrained results in (b).

Appendix C.

Markov chain Monte Carlo

Part of the material presented in this Appendix is adapted from [111].

In Bayesian statistics, one endeavours to test a hypothesis H by combining prior information I with the likelihood of a dataset D, using Bayes' theorem (Equation 4.3) to give you the posterior probability of that hypothesis. The hypothesis could be a specific value of a parameter, which you would estimate by calculating the likelihood of that parameter value with the data set you have collected, and also the likelihood according to any prior knowledge. Prior knowledge could be a previous measurement, in which you constrain the value to a best fit value with a corresponding Gaussian error. This single valued hypothesis would result in a single valued posterior probability. However, we wish to test many hypotheses, in order to find the most probable hypothesis. This could be done by doing a sequential grid search by varying the proposed value of a parameter, constructing the MC prediction according to your PDF, testing this against your data, computing the posterior probability, and then repeat for the next point. You would then expect to see varying probabilities for your various proposed values of parameters, and you would say that the best fit of that parameter is the value that has the highest posterior probability. This may be OK in one or two dimensions, but you can imagine extending this process into higher dimensions would result in a computational nightmare. How can one approximate a continuous multi-dimensional probability distribution efficiently?

Markov chain Monte Carlo (MCMC) is a technique that is particularly useful for computing high dimensional integrals. The concept is to use a semi-random walk through parameter space sampling the likelihood function, creating a collection of discrete steps whose density is proportional to that of the posterior distribution. This negates the need

to sample areas of low probability, because the MCMC naturally tends towards the high probability regions.

C.1. Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm [112] is used to simulate a Markov chain by generating new steps that depend on the current step according to the transition probability $p(X_{t+1}|X_t)$.

The MH algorithm works with 2 steps. The first is to propose a candidate step by drawing from a proposal function $q(Y|X_t)$, where Y is the proposed step and X_t is the current step. Second, this step must be evaluated and either accepted as the new step in the chain by setting $X_{t+1} = Y$ or rejected, keeping the current step $X_{t+1} = X_t$. This is done according to the MH algorithm by calculating the Metropolis ratio:

$$r = \frac{p(Y|D,I)q(X_t|Y)}{p(X_t|D,I)q(Y|X_t)}. \tag{C.1}$$

If $r \geq 1$ then the step is accepted. If $r < 1$ then a random sample U is generated from a uniform distribution in the range 0 to 1, and if $r \geq U$ then the step is accepted. More formally, this operation is called the acceptance probability α and is defined as

$$\alpha(X_t, Y) = \min\left(1, \frac{p(Y|D,I)q(X_t|Y)}{p(X_t|D,I)q(Y|X_t)}\right). \tag{C.2}$$

When combined with the proposal function, we arrive at the definition of the transition probability

$$p(X_{t+1}|X_t) = q(X_{t+1}|X_t)\alpha(X_t, X_{t+1}) \tag{C.3}$$

which is time dependent, and has the property of generating samples whos probability density is equal to that of the posterior $p(X|D,I)$ target distribution, once a burn-in period has been removed from the chain.

It is interesting to note that the original Metropolis algorithm [113] was adapted by W. K. Hastings [112] in 1970 to a more general form that works with non-symmetric proposal

functions. When the proposal function is symmetric, the second term in Equation C.1 becomes 1, and the original Metropolis algorithm is restored.

Regardless of the proposal function chosen, the MCMC will always construct the posterior distribution, a process also referred to as converging to a stationary distribution. For this to happen, the chain must have 3 properties:

- irreducible: From all possible initial states, the Markov chain must be able to reach all other potential states with positive probability.
- aperiodic: A property that stops the chain from moving periodically back-and-forth between two states
- positive recurrent: When the stationary distribution has been reached, then all subsequent steps are samples from the same target distribution.

C.2. Step-size

The MCMC generates a new step by sampling a proposal function for each parameter. Typically the proposal function is a gaussian distribution, the σ of which can correspond to the stepsize of the parameter. An increase of the width of the gaussian increases the 1 dimensional range that the parameter can step.

The proposal function dictates how fast the MCMC will converge to a stationary distribution; the posterior distribution is completely independent of the proposal function. Perhaps the most troublesome feature of the MCMC technique is that tuning the step size for each parameter is notoriously difficult. One must take into account the correlations between parameters, along with increasing dimensionality. A discussion of the techniques to determine the correct step size for a gaussian model can be found in [114]. The authors determine, via a spectral analysis, the relation between the ideal step size and the dimensionality of a gaussian model

$$\sigma_T \approx \sigma_0 \frac{2.4}{\sqrt{D}} \tag{C.4}$$

where σ_T is the tuned step size, σ_0 is the true width of the posterior distribution and D is the number of dimensions of parameter space. The value 2.4 was determined experimentally from the top plot of Figure C.1, where the most efficient scaling value of

the step size (expressed as $\sqrt{D}(\frac{\sigma_T}{\sigma_0})$) is 2.4. The bottom plot of Figure C.1 shows the efficiency as a function of MCMC acceptance, where it is found that the ideal acceptance rate for a 1 dimensional model is ≈ 0.45, and approaches ≈ 0.25 for higher dimensions. The efficiency of the MCMC is defined [115] as

$$E = \lim_{N \to \infty} \frac{\sigma_0^2/N}{\sigma_{\bar{x}}^2(N)} \tag{C.5}$$

where σ_0^2 is the variance of the true distribution, $\sigma_{\bar{x}}^2(N)$ is the variance of a chain of length N. In the long chain limit, E^{-1} gives the factor by which the chain is longer than an optimum MCMC whilst having the same performance.

A simple toy MCMC on a 2 parameter model was built using the R language [116] to demonstrate the effect of step size on the quality of the result. In the MCMC run in Figure C.2, the step size is too small, and the chain is highly correlated. This is because the chain cannot step far enough, and so explores the local density. In contrast, Figure C.3 shows a chain with a step size that is too large. The chain proposes steps which more are frequently not in high probability region, and so the acceptance is poor. Figure C.4 shows a chain with a good step size. The chain has minimum correlations, has good mixing and results in converging to a stationary distribution in a relatively fast time scale, when compared to Figures C.2 and C.3. As previously stated, though the step size may be suboptimal, given enough steps, all 3 examples chains will converge.

C.3. Burn-in

The MCMC typically starts not in the region of high probability, and so the chain must be allowed to explore the space until it enters equilibrium. This burn-in period is discarded, as the steps are correlated and do not represent samples from the posterior distribution. After the burn-in period, the chain meets the 3 requirements to be a Markov process described in Section C.1. An example of a burn-in period is shown in Figure C.5.

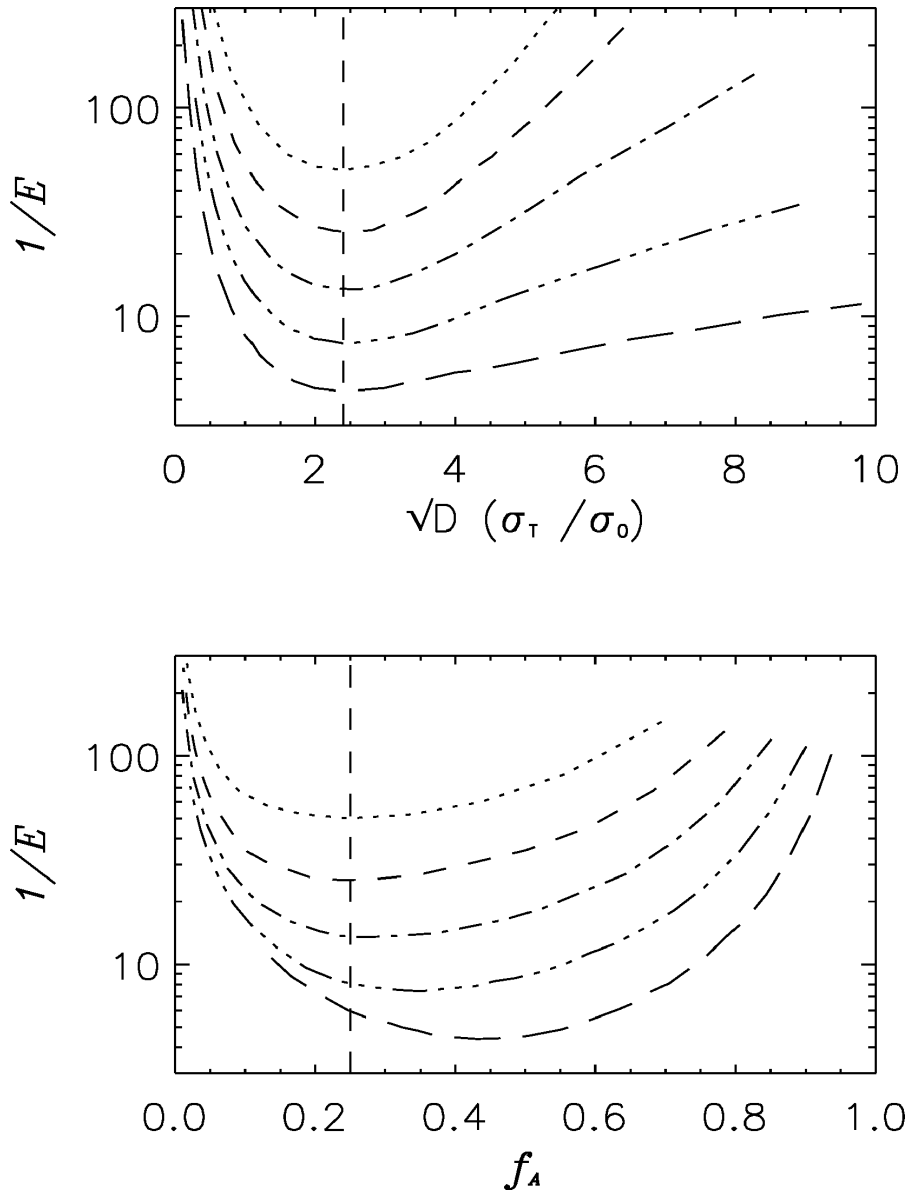


Figure C.1.: The inverse efficiency of the MCMC as a function of step size (top) and acceptance (bottom). The lines on both plots correspond to $D = 1$ (bottom line), 2, 4, 8 and 16 (top line). Figures taken from [114]

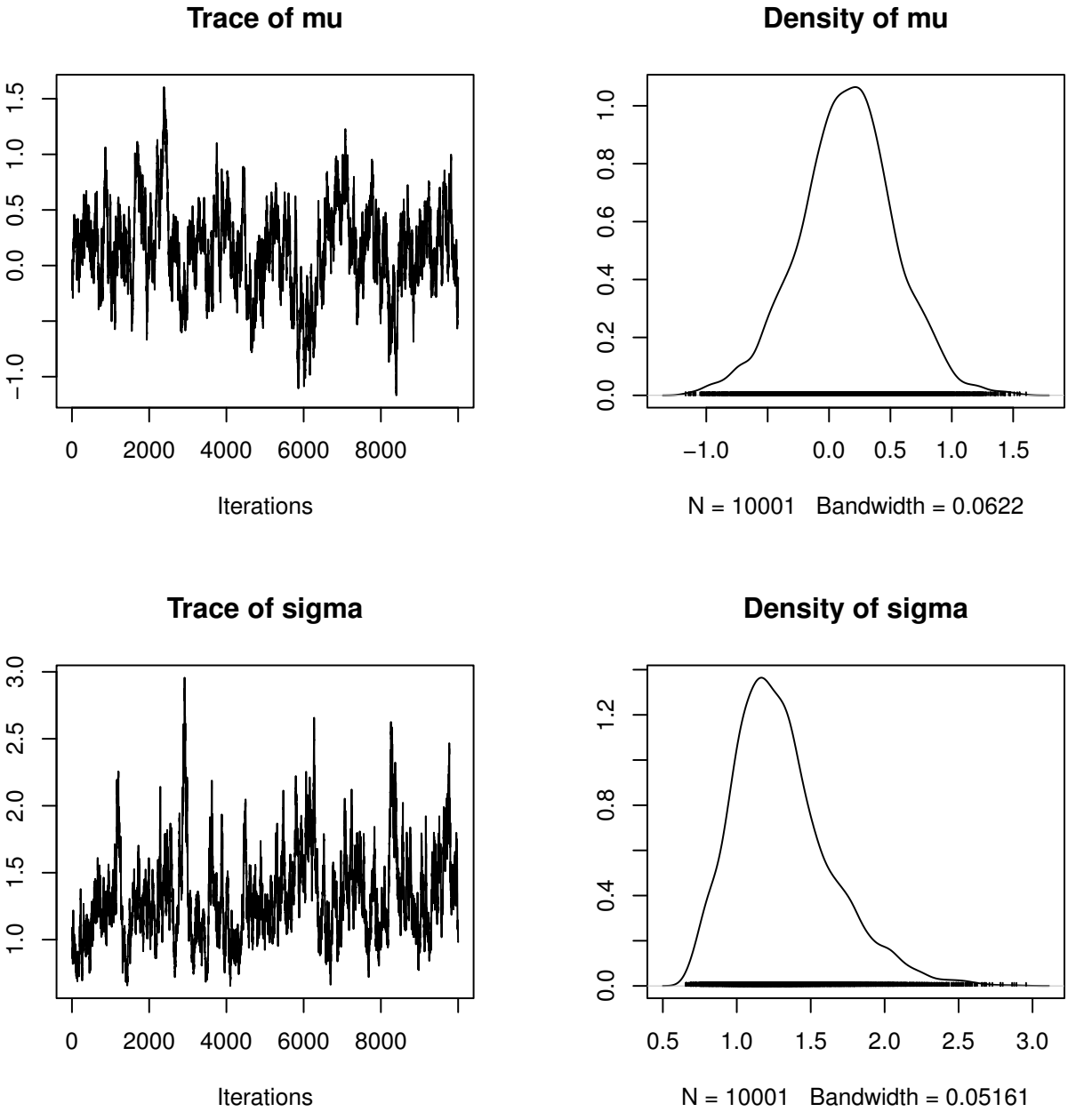


Figure C.2.: Too small stepsize

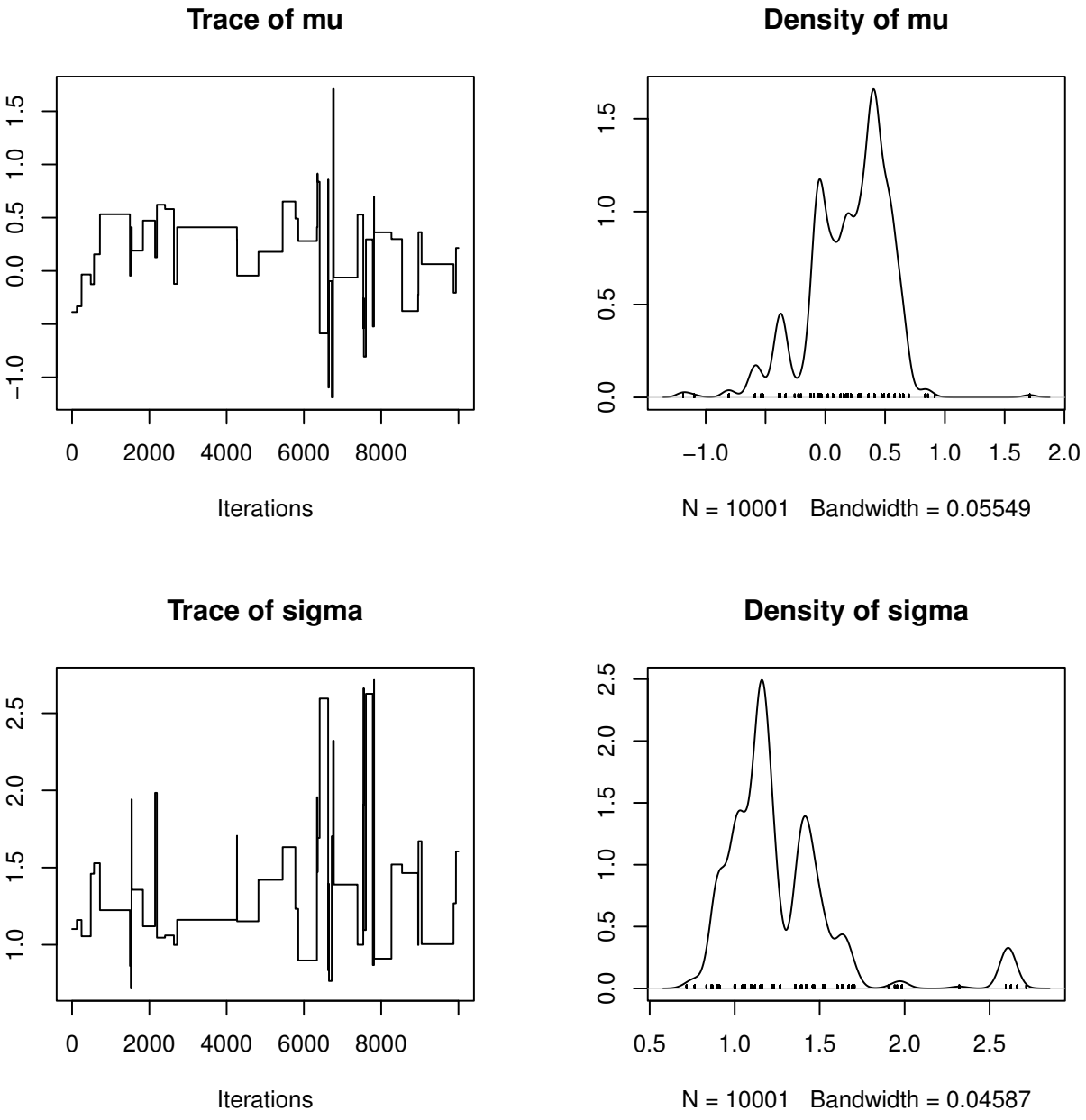


Figure C.3.: Too big stepsize

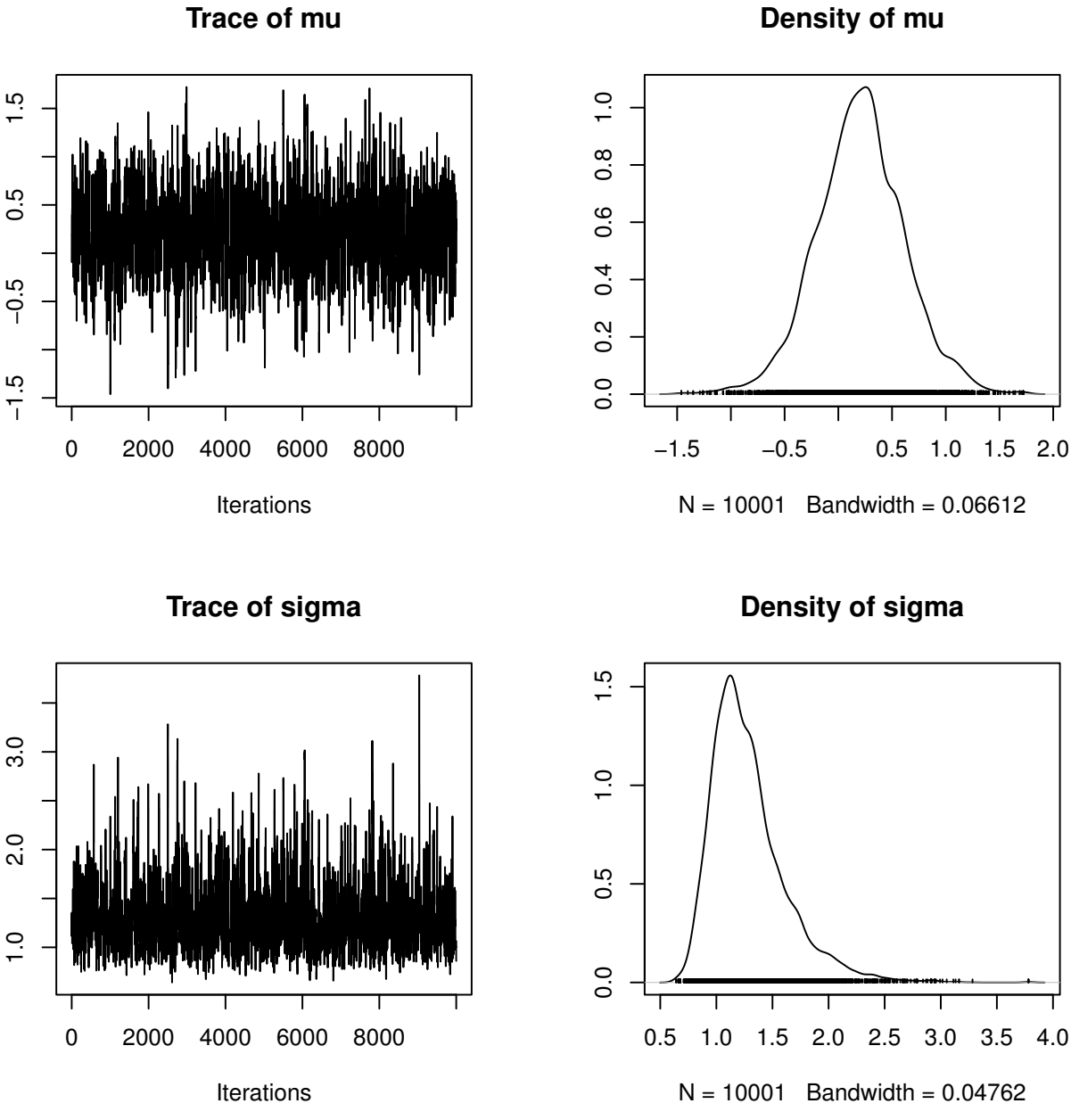


Figure C.4.: Good stepsize

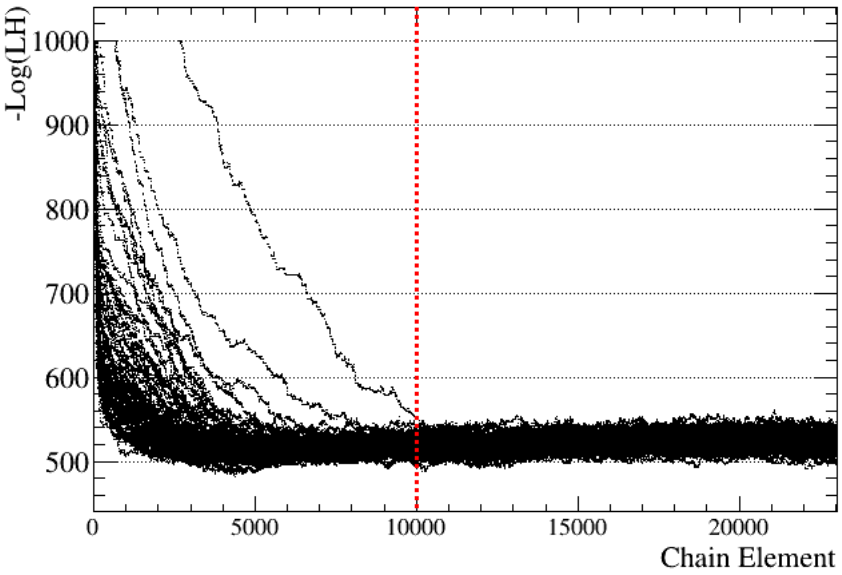


Figure C.5.: The elements of 100 MCMC runs using the same model and data. The starting configuration of each chain is chosen at random from the proposal function centered at the nominal value of each parameter. The different configurations of parameters produce different starting likelihoods, and so each chain has a different burn-in period. For these 100 chains, all of them have reached the region of high probability by 10000 steps.

Appendix D.

The Compute Unified Device Architecture

The material presented here is partially summarised from the CUDA Programming Guide [86]. As the demand for the consumer Graphics Processing Unit (GPU) has risen, modern GPUs have become parallel manycore processors with a large computational

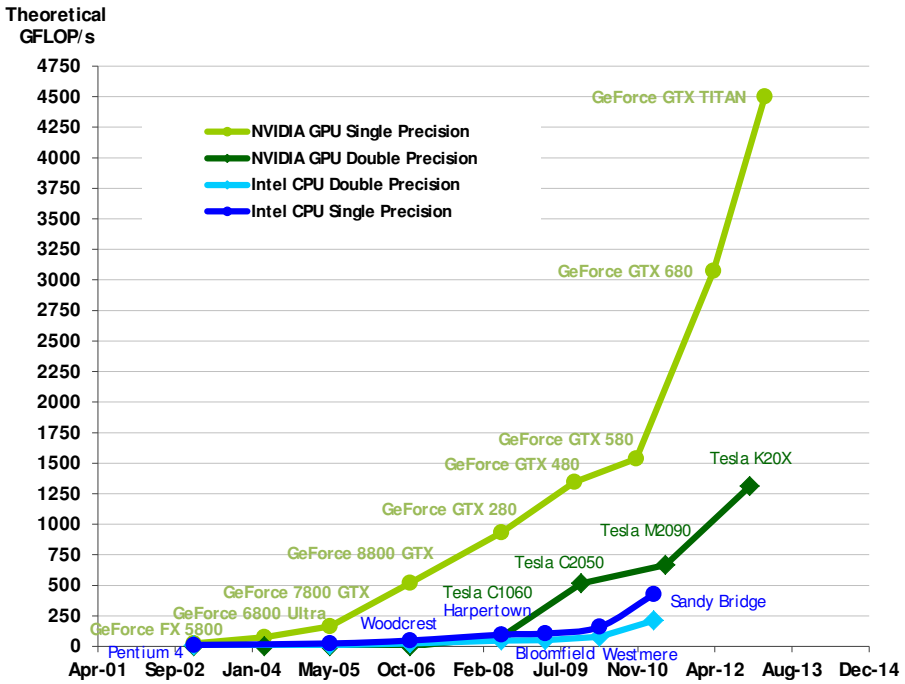


Figure D.1.: Theoretical floating point operations per second of cutting edge hardware from the last decade. CPUs and GPUs are compared, and it can be seen that GPUs have the clear advantage. *GeForce* and *Tesla* are model names for NVIDIA consumer and production products respectively. Figure taken from [86].

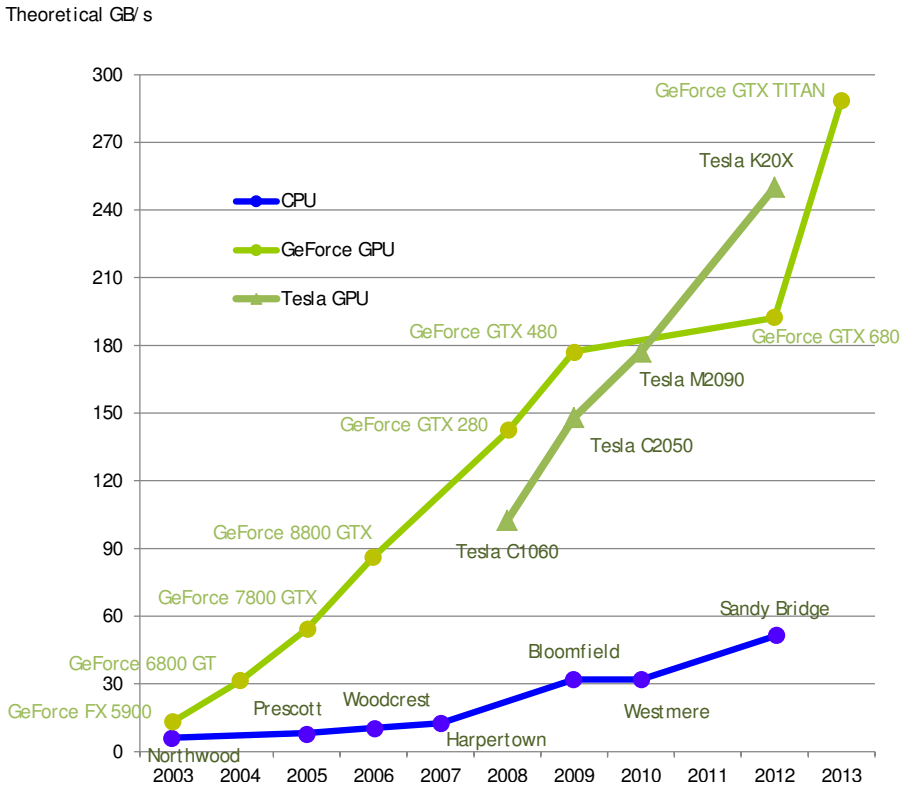


Figure D.2.: Theoretical memory bandwidth (GB/s) of cutting edge hardware from the last decade. Devices are the same as those shown in Figure D.1. Figure taken from [86].

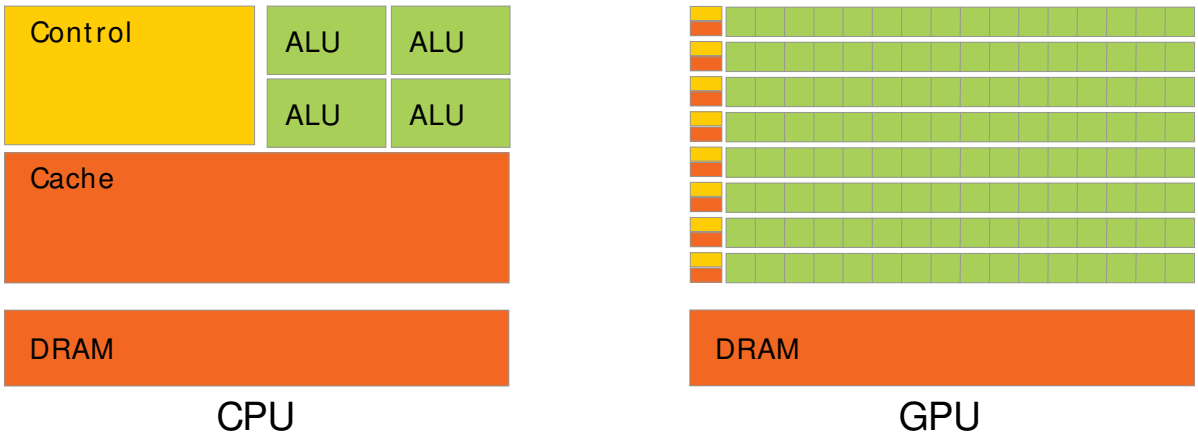


Figure D.3.: A comparison of the physical layout of CPU and GPU transistors. The GPU devotes more transistors to data processing (green) than data caching (red) and flow control (orange). Figure taken from [86].

power and memory bandwidth, as shown in Figures D.1 and D.2. Figure D.1 shows that the GPU has a much higher number of theoretical (giga) floating point operations per second (GFLOP/s) than a CPU, and also a comparable advantage over the CPU in memory bandwidth (Figure D.2). This is because GPUs are specialized devices designed for compute-intensive, highly parallel calculations originally catering to the high-end 3D graphics market. As seen in Figure D.3, the GPU dedicates more transistors to processing rather than data caching and flow control. This makes GPUs well-equipped to solve data-parallel problems where the same program is executed on multiple data in parallel. This makes a GPU a single instruction multiple data (SIMD) device, which means the need for advanced flow control is less. Memory access latency associated with a more basic flow control can be hidden by achieving a high arithmetic intensity, i.e. the ratio of computation to memory operations. In data-parallel processing, individual data elements are mapped to parallel processing threads, which means that applications that require the processing of large data sets can use data-parallel processing to achieve significant speed improvements.

D.1. Compute Unified Device Architecture

The Compute Unified Device Architecture (CUDA) is a parallel computing platform and programming model released by NVIDIA for use on their GPU hardware [86]. CUDA enables the GPU hardware traditionally designed for graphics-based usage to be exposed

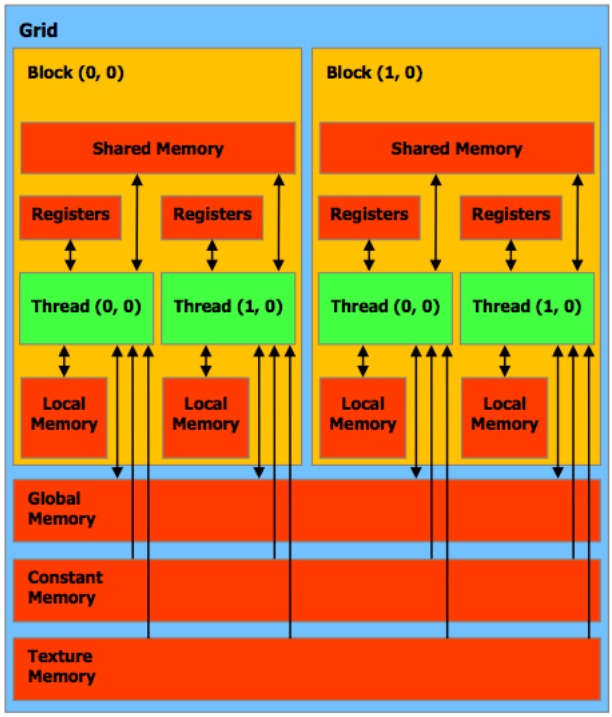


Figure D.4.: Figure taken from [117]

for non-graphics computation via an extension of the C language. Other interfaces and application programming interfaces exist, but are not considered for this appendix.

CUDA is a heterogeneous programming model that operates using a host (CPU) and device (GPU). Explicity management of the GPU memory is required, as users must allocate, copy to and from, and free GPU memory.

D.2. GPU Architecture

In CUDA, each processor core on the GPU is called a thread. The thread maps to a single data element in GPU memory. A kernel is the program that operates on all data elements. Threads are grouped into blocks, which are an abstraction for programming purposes. All threads in a single block are executed on the same stream processor (SM), which allows them to communicate via the shared memory bank. A group of blocks form a grid. This abstraction is shown in Figure D.4. Also shown is the memory hierarchy of a CUDA device. Individual threads have access to registers. These are typically fast but small caches on a per thread basis. A block, as previous mentioned, can access

a shared memory bank (Labelled *local memory* in Figure D.4), which is visible to all threads within the block. This is a fast cache that can be used to create complex parallel algorithms where the threads in a block work together, communicating via the shared memory. However, the programmer must be careful as shared memory is susceptible to race conditions and bank conflicts. All types of memory mentioned so far exist on the GPU die. The remaining memory types are off-chip RAM that is slower but has a much higher capacity. The video RAM of the GPU is partitioned into 3 types: global memory, constant memory and texture memory, which are accessible by all threads on the GPU. Global memory is the most commonly used, and is where data is copied to from the host. Typically global memory is on the order of gigabytes, however its access times are restricted to clock speeds on the order of 1 GHz. Constant and texture memory are small areas of reserved memory on the video RAM that serve as cached memory for specific uses. Constant memory is read-only, and performs faster for accessing data items that stay the same throughout the program. Texture memory is derived from the use of GPUs for graphical applications, where the memory is read-only and locally cached; adjacent data elements in memory are accessed faster than if it were randomly accessed.

Despite this complicated memory hierarchy, a simple parallel application gaining orders of magnitude speed improvement can be programmed using only global memory. The exploitation of the other types of memory can achieve even more performance, however more advanced parallel algorithms must be employed.

D.3. An Example Program

An example calculation of squaring an array of numbers is now shown. The complete code and compilation instructions are given so that this can be reproduced.

Listing D.1 contains the host code that contains the main function. This involves instantiating an array of numbers. Listing D.3 contains code that handles the copying to and from the GPU memory, along with executing the kernel function. Finally, Listing D.2 contains the kernel function that is executed on each element of the array of numbers, which squares each element in parallel, overwriting the input with the squared value. The original array is copied back to the host memory where it is printed to the screen.

Listing D.1: "main.cpp" – Host code.

```
#include <cstdlib>
```

```
#include <iostream>

using namespace std;

// define the external function to be compiled separately
extern "C" void runTest(float*);

int main()
{
        // declare host memory
        float *array_host;

        // allocation of host memory, an array of 32 floats
        array_host = (float *)malloc(32*sizeof(float));

        // initialise array with the index
        for (int i=0;i<32;i++) {array_host[i] = (float)i;}

        // iterate over the initialised array and print its contents
        cout << "Input array: " << endl;

        for (int i=0;i<32;i++) {cout << array_host[i] << ", ";}
        cout << endl;

        // run CUDA function, pass host pointer to function
        runTest(array_host);

        // iterate over the result and print the answer
        cout << "Output array: " << endl;

        for (int i=0;i<32;i++) {cout << array_host[i] << ", ";}
        cout << endl;

}
```

Listing D.2: "kernel.cu" – Kernel code.

```
__global__ void squared(float *input)
{
```

```
    // create thread id from the dimension of the block, the block id and the
        thread ID
        int idx = (blockIdx.x * blockDim.x + threadIdx.x);

        // square the element of the array corresponding to this thread id
        input[idx] = input[idx] * input[idx];
}
```

Listing D.3: "cuda.cu" – CUDA code.

```
#include "kernel.cu"

extern "C" __host__ void runTest(float *host_array)
{
        int num_blocks = 1;
        int block_size = 32;

        // create a pointer to device memory
        float *device_array;

        // specify size of array
        size_t size = 32 * sizeof(float);

        // CUDA function to allocate memory on the GPU
        cudaMalloc((void **) &device_array, size);

        // copy the array to the device
        cudaMemcpy(device_array, host_array, size, cudaMemcpyHostToDevice);

        // execute kernel over a grid of thread blocks, number of threads =
            num_blocks*block_size
        squared<<<num_blocks,block_size>>>(device_array);

        // copy results from device back to the host
        cudaMemcpy(host_array, device_array, size, cudaMemcpyDeviceToHost);

        // free device memory
        cudaFree(device_array);
}
```

To compile the example, the NVIDIA compiler `nvcc` must be used to compile device code into object code:

```
nvcc -c cuda.cu.o cuda.cu
```
(D.1)

Next, compile the host code into object code:

```
g++ -Wall -c -o main.o main.cpp
```
(D.2)

Finally, link both sets of object code into an executable:

```
g++ -o square cuda.o main.o -lcudart
```
(D.3)

It is assumed that the CUDA toolkit is installed and already in the users path.

Bibliography

[1] R. G. Calland, Construction of the Barrel Electromagnetic Calorimeter for the T2K Experiment, Master's thesis, University of Liverpool, 2010.

[2] R. G. Calland, A. C. Kaboth, and D. Payne, Journal of Instrumentation **9**, P04016 (2014).

[3] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).

[4] J. Boger *et al.*, Nuclear Instruments and Methods in Physics Research A **449**, 172 (2000), nucl-ex/9910016.

[5] (Super-Kamiokande Collaboration), Y. Fukuda *et al.*, Phys. Rev. Lett. **81**, 1562 (1998).

[6] C. Giunti and C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics* (Oxford University Press, Oxford, UK, 2007), ISBN 978-0-19-850871-7.

[7] Z. Maki, M. Nakagawa, and S. Sakata, Progress of Theoretical Physics **28**, 870 (1962).

[8] (T2K Collaboration), K. Abe *et al.*, Phys. Rev. Lett. **112**, 061802 (2014).

[9] F. P. An *et al.*, Physical Review Letters **108**, 171803 (2012), 1203.1669.

[10] V. Barger, K. Whisnant, S. Pakvasa, and R. J. N. Phillips, Phys. Rev. D **22**, 2718 (1980).

[11] J. N. Bahcall and C. Peña-Garay, New Journal of Physics **6**, 63 (2004).

[12] Image from, `http://www.sns.ias.edu/~jnb/Papers/Preprints/solarmodels.html`, 2007.

[13] J. N. Bahcall, Phys. Rev. Lett. **12**, 300 (1964).

[14] B. T. Cleveland *et al.*, The Astrophysical Journal , 505+ (1998).

[15] SAGE Collaboration, J. Abdurashitov *et al.*, Phys.Rev. **C60**, 055801 (1999), astro-ph/9907113.

[16] W. Hampel *et al.*, Physics Letters B **447**, 127 (1999).

[17] GNO COLLABORATION, M. Altmann *et al.*, Phys.Lett. **B616**, 174 (2005), hep-ex/0504037.

[18] SAGE, J. N. Abdurashitov *et al.*, Phys. Rev. **C80**, 015807 (2009), 0901.2200.

[19] GALLEX, W. Hampel *et al.*, Phys. Lett. **B447**, 127 (1999).

[20] J. N. Bahcall, S. Basu, and M. Pinsonneault, Phys.Lett. **B433**, 1 (1998), astro-ph/9805135.

[21] K. S. Hirata *et al.*, Phys. Rev. Lett. **63**, 16 (1989).

[22] Y. Fukuda *et al.*, Physical Review Letters **81**, 1158 (1998), hep-ex/9805021.

[23] B. Kayser, ArXiv e-prints (2008), 0804.1497.

[24] KamLAND Collaboration, K. Eguchi *et al.*, Phys. Rev. Lett. **90**, 021802 (2003).

[25] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Science **124**, 103 (1956), http://www.sciencemag.org/content/124/3212/103.full.pdf.

[26] S. Abe *et al.*, Physical Review Letters **100**, 221803 (2008), 0801.4589.

[27] M. H. Ahn *et al.*, Phys. Rev. D**74**, 072003 (2006), hep-ex/0606032.

[28] M. Apollonio *et al.*, European Physical Journal C **27**, 331 (2003), hep-ex/0301017.

[29] MINOS Collaboration, P. Adamson *et al.*, Phys.Rev.Lett. (2014), 1403.0867.

[30] MINOS Collaboration *et al.*, ArXiv e-prints (2014), 1403.0867.

[31] F. P. An *et al.*, Physical Review Letters **108**, 171803 (2012), 1203.1669.

[32] T2K Collaboration, K. Abe *et al.*, (2014), 1403.1532.

[33] W. Winter, Phys. Rev. D**88**, 013013 (2013), 1305.5539.

[34] F. Capozzi *et al.*, ArXiv e-prints (2013), 1312.2878.

[35] T2K Collaboration *et al.*, Nuclear Instruments and Methods in Physics Research

A **659**, 106 (2011), 1106.1238.

[36] Y. Takatomi, The design of t2k off-axis near detector and the basic beam property measurements at t2k, 2012, Slides of a talk given at Lake Louise Winter Institute.

[37] A. Blondel, M. Campanelli, and M. Fechner, CERN Report No. CERN-OPEN-2002-049. CERN-NEUTRINO-FACTORY-NOTE-112. CERN-NUFACT-NOTE-112, 2002 (unpublished).

[38] J. Nieves, F. Sánchez, I. R. Simo, and M. J. V. Vacas, Phys. Rev. D**85**, 113008 (2012), 1204.5404.

[39] J. A. Formaggio and G. P. Zeller, Reviews of Modern Physics **84**, 1307 (2012), 1305.7513.

[40] C. McGrew, First t2k neutrino scillation results, 2011, Slides from BNL Seminar.

[41] S. van der Meer, (1961), CERN-61-07.

[42] D. Renker and E. Lorenz, Journal of Instrumentation **4**, P04004 (2009).

[43] M. Yokoyama *et al.*, Nuclear Instruments and Methods in Physics Research A **610**, 362 (2009), 0807.3147.

[44] M. Yokoyama *et al.*, Nuclear Instruments and Methods in Physics Research A **622**, 567 (2010), 1007.2712.

[45] F. R. Y. Du, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **596**, 396 (2008).

[46] Image from, `http://www-flc.desy.de/pet/`, 2007.

[47] P.-A. Amaudruz *et al.*, Nuclear Instruments and Methods in Physics Research A **696**, 1 (2012), 1204.3666.

[48] T2K ND280 TPC collaboration, ArXiv e-prints (2010), 1012.0865.

[49] J. Beucher, Journal of Instrumentation **5**, P01007 (2010).

[50] D. Allan *et al.*, Journal of Instrumentation **8**, 19P (2013), 1308.3445.

[51] S. Assylbekov *et al.*, Nuclear Instruments and Methods in Physics Research A **686**, 48 (2012), 1111.5030.

[52] S. Aoki *et al.*, Nuclear Instruments and Methods in Physics Research A **698**, 135

(2013), 1206.3553.

[53] K. Abe *et al.*, Nuclear Instruments and Methods in Physics Research A **694**, 211 (2012), 1111.3119.

[54] S. Fukuda *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **501**, 418 (2003).

[55] Image from, `http://www.ps.uci.edu/~tomba/sk/tscan/compare_mu_e/`, 2006.

[56] Y. Hayato, Acta Phys.Polon. **B40**, 2477 (2009).

[57] C. Andreopoulos *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **614**, 87 (2010).

[58] N. Abgrall, ArXiv e-prints (2010), 1005.3692.

[59] K. Abe *et al.*, Phys. Rev. D**87**, 019902 (2013), 1211.0469.

[60] A. Fasso' *et al.*, ArXiv Physics e-prints (2003), physics/0306162.

[61] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini.

[62] C. Zeitnitz and T. Gabriel, Nucl.Instrum.Meth. **A349**, 106 (1994).

[63] GEANT4, S. Agostinelli *et al.*, Nucl. Instrum. Meth. **A506**, 250 (2003).

[64] C. H. L. Smith, Phys. Rep. **3**, 261 (1972).

[65] D. Rein and L. Sehgal, *Neutrino excitation of Baryon resonances and single pion production*PITHA (Physikal. Inst., 1980).

[66] D. Rein and L. M. Sehgal, Nuclear Physics B **223**, 29 (1983).

[67] M. Glück, E. Reya, and A. Vogt, European Physical Journal C **5**, 461 (1998), hep-ph/9806404.

[68] R. D. Woods and D. S. Saxon, Phys. Rev. **95**, 577 (1954).

[69] D. Allan *et al.*, Journal of Instrumentation **8**, 19P (2013), 1308.3445.

[70] S. Boyd, Cross section measurements in the t2k nd280 detector.

[71] M. Yokoyama, Application of hamamatsu mppc to t2k near neutrino detectors, 2008.

[72] L. Bellantoni and P. Rubinov, D0 note 4845 .

[73] A. Vacheret, S. Greenwood, N. Noy, M. Raymond, and W. A., Proc. Nuclear Science Symp. Conf (2007).

[74] P. de Perio *et al.*, T2K Report No. TN-106, 2012 (unpublished), `www.t2k.org/docs/technotes/106`.

[75] A. Krishnamoorthy and D. Menon, ArXiv e-prints (2011), 1111.4144.

[76] K. Sakashita *et al.*, T2K Report No. TN-114, 2012 (unpublished), `www.t2k.org/docs/technotes/114`.

[77] P. de Perio *et al.*, T2K Report No. TN-113, 2012 (unpublished), `www.t2k.org/docs/technotes/113`.

[78] R. Brun *et al.*, ROOT web page, `http://root.cern.ch/`, 2001.

[79] C. de Boor, *A Practical Guide to Splines* (Springer-Verlag, 1978).

[80] K. Connolly, Regarding splines, 2013, Talk given at September T2K Collaboration Meeting (internal).

[81] CERN Report No., , 2000 (unpublished), `http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html`.

[82] P. Mills, International Journal of Remote Sensing **32 (21)**, 6109 (2011).

[83] D. DeTurck, Math 480 course notes, `http://www.math.upenn.edu/~deturck/m480/m480lec3/m480lec3.html`, 1996.

[84] M. hui Chen and Q. man Shao, Journal of Computational and Graphical Statistics **8**, 69 (1998).

[85] H. Jeffreys, *Theory of Probability*, Third ed. (Oxford, Oxford, England, 1961).

[86] NVIDIA Corporation, *NVIDIA CUDA C Programming Guide*, 2013.

[87] Khronos OpenCL Working Group, *The OpenCL Specification, version 1.0.29*, 2008.

[88] R. Wendell, Prob3++ software for computing three flavor neutrino oscillation probabilities, `http://www.phy.duke.edu/~raw22/public/Prob3++/`, 2012.

[89] OpenMP Architecture Review Board, OpenMP application program interface version 3.0, `http://www.openmp.org/mp-documents/spec30.pdf`, 2008.

[90] R. M. Stallman and G. DeveloperCommunity, *Using The Gnu Compiler Collection: A Gnu Manual For Gcc Version 4.3.3* (CreateSpace, Paramount, CA, 2009).

[91] P. Pomorski, Lecture, University of Waterloo (2013).

[92] N. Whitehead and A. Fit-florea, Precision & performance: Floating point and ieee 754 compliance for nvidia gpus, 2011.

[93] A. Caldwell, D. Kollár, and K. Kröninger, Computer Physics Communications **180**, 2197 (2009), 0808.2552.

[94] F. P. An *et al.*, Physical Review Letters **108**, 171803 (2012), 1203.1669.

[95] J. Hignight *et al.*, T2K Report No. TN-148, 2013 (unpublished), `www.t2k.org/docs/technotes/148`.

[96] S. Tobayam, Recent updates on fitqun, a new event reconstruction algorithm for water cherenkov detectors, 2013, Slides from talk given at 2nd Open Meeting for the Hyper-Kamiokande Project.

[97] N. Abgrall *et al.*, T2K Report No. TN-99, 2011 (unpublished), `www.t2k.org/docs/technotes/099`.

[98] P. de Perio *et al.*, T2K Report No. TN-108, 2012 (unpublished), `www.t2k.org/docs/technotes/108`.

[99] The MiniBooNE Collaboration, A. A. Aguilar-Arevalo *et al.*, Phys. Rev. D **81**, 013005 (2010).

[100] C. Bojechko *et al.*, T2K Report No. TN-152, 2013 (unpublished), `www.t2k.org/docs/technotes/152`.

[101] T2K Collaboration, K. Abe *et al.*, Phys. Rev. D **88**, 032002 (2013).

[102] P. de Perio *et al.*, T2K Report No. TN-186, 2014 (unpublished), `www.t2k.org/docs/technotes/186`.

[103] Particle Data Group, J. Beringer *et al.*, Phys. Rev. D **86**, 010001 (2012).

[104] J. Beringer and others (PDG), Phys.Rev. **D86**, 010001 (2012), `http://pdg.lbl.gov)`.

[105] L. Dumortier and L. Lyons, CDF Report No. CDF/ANAL/PUBLIC/5776, 2002 (unpublished), `www.t2k.org/docs/technotes/166`.

[106] M. Hartz, A. Kaboth, and K. Mahn, T2K Report No. TN-166, 2013 (unpublished), `www.t2k.org/docs/technotes/166`.

[107] K. Hagiwara, N. Okamura, and K.-I. Senda, Journal of High Energy Physics **9**, 82 (2011), 1107.5857.

[108] A. Gelman, X. Meng, and H. Stern, Statistica Sinica **6**, 733 (1996).

[109] MINOS Collaboration, P. Adamson *et al.*, Phys. Rev. Lett. **110**, 171801 (2013).

[110] A. Himmel and for the Super-Kamiokande Collaboration, ArXiv e-prints (2013), 1310.6677.

[111] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, New York, NY, USA, 2005).

[112] W. K. Hastings, Biometrika **57**, 97 (1970).

[113] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Journal of Chemical Physics **21**, 1087 (1953).

[114] J. Dunkley, M. Bucher, P. G. Ferreira, K. Moodley, and C. Skordis, Mon. Not. Roy. Astron. Soc. **356**, 925 (2005), astro-ph/0405462.

[115] Department of Computer Science, University of Toronto, Toronto Report No., , 1993 (unpublished).

[116] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.

[117] T. Warburton, Rocky mountain mathematics consortium summer 2008: Parallel numerical methods forpartial differential equations, 2008.